OM protein - protein search, using sw model

Copyright (c) 1993 - 2002 Compugen Ltd.

GenCore version 5.1.3

Run on: October 19, 2002, 07:41:36 ; Search time 44 Seconds
                 (without alignments)
                 1000.203 Million cell updates/sec

Title: US-09-807-459-2
Perfect score: 2359
Sequence: 1 MAPSDSVGDVTKTLLAASES.......DPSKALIRKVSTEADNLLEK 458

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR.71:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

Result   %
No.   Score   Query Match Length DB   DB ID   Description

  1   827.5   35.1   480   1   A45614   merozoite surface
  2   826.5   35.0   565   2   S27778   merozoite surface
  3   823   34.9   442   2   B45561   merozoite surface
  4   813.5   34.5   480   2   D45561   merozoite surface
  5   790   33.5   442   2   C45561   merozoite surface
  6   770.5   32.7   456   2   C48572   merozoite surface
  7   632.5   26.8   321   2   A48572   rhoptry protein ho
  8   366   15.5   192   2   B48572   rhoptry protein ho
  9   143.5   6.1   441   2   T28411   rhoptry protein ho
 10   143.5   5.7   441   2   T28411   ORF MSV250 hypothe
 11   134   5.6   1098   2   B70232   hypothetical prote
 12   133   5.6   624   2   PC6003   surface membrane p
 13   127   5.4   607   2   S48326   hypothetical prote
 14   125   5.3   1830   2   E82909   conserved hypothet
 15   122.5   5.2   460   2   E70132   hypothetical prote
 16   122   5.2   1829   2   S35027   cytotoxin RTX homo
 17   121   5.1   566   1   HMIVUR   hemagglutinin homo
 18   118.5   5.0   2441   2   D71623   hemagglutinin prec
 19   117.5   5.0   447   2   A72358   erythrocyte membra
 20   117   5.0   986   2   H90565   conserved hypothet
 21   117   5.0   1038   1   B42641   restriction modifi
 22   117   5.0   1829   2   B81086   kinesin-related pr
 23   117   5.0   2166   2   G70163   iron-regulated pro
 24   116.5   4.9   457   2   G97085   hypothetical prote
 25   116   4.9   457   2   G82925   sensory transducti
 26   116   4.9   1250   2   E81339   hypothetical prote
 27   115   4.9   912   2   H71931   probable restricti
 28   115   4.9   986   2   S49394   hypothetical prote
 29   115   4.9   985   2   H71931   hypothetical prote
 30   114   4.8   480   1   A45614   merozoite surface
 31   114   4.8   652   2   F90561   hypothetical prote
 32   113.5   4.8   2116   2   A266655 myosin heavy chain
 33   113.5   4.8   2364   2   A56577   microtubule-associ
 34   112.5   4.8   1314   2   A85176   hypothetical colle
 35   112   4.7   1957   2   T38077   hypothetical prote
 36   111.5   4.7   566   2   T50382   probable spindle p
 37   111   4.7   842   2   G90576   heat shock atp-dep
 38   110.5   4.7   812   2   A46417   NIP1 protein - yea
 39   110.5   4.7   1188   2   T20033   nuclear fusion pro
 40   110   4.7   566   1   HMIV   hemagglutinin prec
 41   110   4.7   577   2   T43207   hypothetical prote
 42   110   4.7   1271   2   D64237   bacitracin synthet
 43   109.5   4.6   1076   2   B84687   probable kinesin-1
 44   109.5   4.6   1939   2   T18372   repeat organellar
 45   109   4.6   420   2   B72386   hypothetical prote

                              ALIGNMENTS

RESULT 1
A45614
merozoite surface antigen p58 - Babesia bigemina
C;Species: Babesia bigemina
C;Date: 22-Apr-1993 #sequence_revision 02-Jun-1994 #text_change 16-Jul-1999
C;Accession: A45614; A45561; S27775
R;Mishra, V. S.; Stephens, E. B.; Dame, J. B.; Perryman, L. E.; McGuire, T. C.; McElwain,
Mol. Biochem. Parasitol. 47, 207-212, 1991
A;Title: Immunogenicity and sequence analysis of recombinant p58: a neutralization-se
A;Reference number: A45614; MUID:92045553
A;Accession: A45614
A;Molecule type: DNA
A;Residues: 1-480 <MIS1>
A;Cross-references: EMBL:M60878; NID:g155860; PIDN:AAA65583.1; PID:g155861
A;Experimental source: merozoite
A;Note: sequence extracted from NCBI backbone (NCBIN:65037, NCBIP:65038)
R;Mishra, V. S.; McElwain, T. F.; Dame, J. B.; Stephens, E.B.
Mol. Biochem. Parasitol. 53, 149-158, 1992
A;Title: Isolation, sequence and differential expression of the p58 gene family of Ba
A;Reference number: A45561; MUID:92365724
A;Accession: A45561
A;Molecule type: DNA
A;Residues: 1-281, 'G', 283-480 <MIS2>
A;Note: sequence extracted from NCBI backbone (NCBIN:111161, NCBIP:111164)
A;Gene: p58
A;Genetics:
A;Superfamily: merozoite surface antigen p58
C;Keywords: surface antigen

Query Match   35.1%; Score 827.5; DB 1; Length 480;
Best Local Similarity   41.8%; Pred. No. 1.6e-46;
Matches 181; Conservative 75; Mismatches 140; Indels 37; Gaps 12;

Qy   4 SDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSASY 63
          ::   :||: |:: || | :||:|:: | :|:::: |:|||||:|:| ::: :
Db   33 AEVVGDVSKTLLEANEVVNAEMEATQVNKDMQSOLSNVKETIVGEVCERVAGNSTCGESV 92

Qy   64 SAYMSRCAKODCLTLQSLKYPLEARKYQPLTLPDPYQLERAAFILFKESDANPANSTEKRFW 123
          |||:|: || :||:|: || :|||| |:||: ||:||||:::||| ||: ||| || ||
Db   93 IAYVNRCDEGDCLTLDSM------KYKPLSLPNPYQLDDAAFMLFRESDSNPAKNEVKRFW 146

Qy   124 MRFRGRKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEFGAS 183
          ||:|: ||:|| ||| |: :| ::|:| :|| :|| :|||||:: ||:|| :|| :| |:
Db   147 MRSR--SSHGDYHHFVSLLKKNVVRDPESNDVENFASQYFYMTTLYYKTYLTVDFTAAK 204

Qy   184 FFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLOHITSSSXKDYMDYQIPALP 243
          |||| || || |:|: ||||||::|:: ::|: :|::||: |:: ::||| |:: : :|
Db   205 FFNKLAFTTRLFGFGCIQKALKRLVRSNLPVDLGT-HPEATIREIASGYGEYMMTQVPAMT 263

Qy   244 KFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREPSK- 302
          ||| |||:||||| |||: |: : ||| :| |:|||::|| :||| ::|||| ||
Db   264 SFAERFSKMATKTLLVTVSDYVHLPAYKRWYRKFEFIVN-FFTDPAKLIMKHVSQPVKT 322

Thu Oct 24 14:26:30 2002   us-09-807-459-2.rpr   Page 2

```
Qy    303  ALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKI------G  356
               : :|| : ||       ::  :|  |  :   |  :   |    :|      |
Db    323  AYTKLVPEEHRQAIRNVVGQSTKHIAN---GVRDLSRMIKE----PSQQIIREKLPHYLSKA  377

Qy    357  QGTVDFINNEIRD--PSKALIRKVSTGAEDLFENKIGQGTVDFINNEIRDPSKALIRKVY  414
               ::  :|  :       :  :   :  |                              :
Db    378  KGAVEHVVKVKSVVPIK---QKDQPSEAAVEETVPSG--DSAETEFEVPEEQYVDAVT  432

Qy    415  T--------EADD  419
           |        ||||
Db    433  TQEVNSEKVDADD  445
```

RESULT 2
S27778
merozoite surface antigen, 60K - Babesia bovis
C;Species: Babesia bovis
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C;Accession: S27778
R;Suarez, C.E.; Palmer, G.H.; Jasmer, D.P.; Hines, S.A.; Perryman, L.E.; McElwain, T.F.
A;Description: Characterization of the gene encoding a 60 kilodalton Babesia bovis meroz
submitted to the EMBL Data Library, February 1991
A;Reference number: S27778
A;Molecule type: mRNA
A;Accession: S27778
A;Residues: 1-565 <SUA>
A;Cross-references: EMBL:M38218; NID:g155883; PID:g155884
C;Superfamily: merozoite surface antigen p58
C;Keywords: surface antigen

Query Match   35.0%;   Score 826.5;   DB 2;   Length 565;
Best Local Similarity 35.5%;   Pred. No. 2.3e-46;
Matches 178;   Conservative 79;   Mismatches 195;   Indels 49;   Gaps 6;

```
Qy      1  MAPSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCS   60
Db     31  LAPAEVVGDLTGTLTETADTLMTLRDHMHNITKDMKHVLSNGREQIVNDVCSNAPEDSNCR   90

Qy     61  ASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAARFLFKESDANPANSTEK  120
Db     91  EVNNNYADRCEMYGCFTIDNVKPLYQEYQPLSLPNPYQLDAAERLFKESASNPAKNSVK  150

Qy    121  RFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYKTYNVDEF  180
Db    151  REWLRFRNGANHGDYHYFVTGLLNNNVHEEGTIGTDVEYLVNKVLYMATMNYKTYLTVNSM  210

Qy    181  GASFFNKLSFTTIGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIP  240
Db    211  NAKFFNRFSFTTKIFSRRIRQTLSDIIRWNVPEDF-EERSIERITQLTSSYEDYMLTQIP  269

Qy    241  ALPKFAKRRFSLMVVQRLLATVAGYVDTPWYKKWVMKLKNFMVNRVFIPTKKF------  292
Db    270  TLSKFARRYADMVKKVLLGSLTSYEVEAPWYKRWIKKFRDFFSSKNVTQPTRKFIEDTNEVT  329

Qy    293  ------FNKEIREPSKALKEKVSTDTKDLFENKIGQGTVD  326
Db    330  KNYLKANVAEPTKFMQDTHEKTRGYLKENVAEPTKTFFKEAPQVTKHFFDENIGQPTKE  389

Qy    327  FFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLF  386
Db    390  FFREAPQATKHFLDENIGQPTKEFF-REAPQATKHFLGENIAQPTKEFFKRDVPQVTKKVI  448

Qy    387  ENKIGQGTVDF---------INNEIRDPSKALIRKVYTEADDLFENKIGQGTVDFINK  435
Db    449  TENIAQPTKEFRREVPHATMKVLNENIAQPAKEIIHEEFGTGAKN-FISAAHEGTKQFLNE  507

Qy    436  EIRDPSKALIR-KVSTEADNL  455
Db    508  TVGQPTKEFLNGALETTKDAL  528
```

RESULT 3

B45561
merozoite surface antigen (clone 7) - Babesia bigemina
C;Species: Babesia bigemina
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 16-Feb-1997
C;Accession: B45561
R;Mishra, V.S.; McElwain, T.F.; Dame, J.B.; Stephens, E.B.
A;Title: Isolation, sequence and differential expression of the p58 gene family of Ba
Mol. Biochem. Parasitol. 53, 149-158, 1992
A;Reference number: A45561; MUID:92365724
A;Molecule type: DNA
A;Accession: B45561
A;Status: preliminary
A;Residues: 1-442 <MIS>
A;Cross-references: GB:M85184
C;Superfamily: merozoite surface antigen p58
C;Keywords: surface antigen

Query Match   34.9%;   Score 823;   DB 2;   Length 442;
Best Local Similarity 42.3%;   Pred. No. 2.9e-46;
Matches 174;   Conservative 78;   Mismatches 129;   Indels 30;   Gaps 10;

```
Qy      4  SDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSASV   63
Db     33  AEVVGDVSKTLLEANEVVNAEMEATQVNKDMQSQLSNVKETIVGEVCEKVAGNSTCGGESV   92

Qy     64  SAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEKRFW  123
Db     93  IAYVNRCDEGDCLTLDSM----KYKPLSLPNPYQLDAAFMLFRESDSNPAKNEVKRFW  146

Qy    124  MRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYKTYNVDEFGAS  183
Db    147  MRSR--SSHGDYHHFVVSLLKKNVVRDPESNDVENFASQYFYMTTLYKTYLTVDFTAAK  204

Qy    184  FFNKLSFTTIGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALP  243
Db    205  FFNKLAFTTRLPGFGIQKALKRLVRSNLPVDLGT-HPEATIRBIASGYGEYMMTQVPAMT  263

Qy    244  KFPAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKF-FNKEIREPSK-  302
Db    264  SFAERFSKMATKTLLVTVSDYVHLPAYKRWYRKFKEFIVN-FFTDPAKLIMKHVSQPVKT  322

Qy    303  ALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKI------G  356
Db    323  AYTKLVPEEHRQAIRNVVGQSTKHIAN--GVRDLARMIKE----PSQQIIREKLPHYLSKA  377

Qy    357  QGTVDFINNEIRDPSKALIRKVSTGAEDLFEN------KIGQGTVDFINNE  401
Db    378  KGAVEHVVDKVK--SKTFKKRAGESSEESYRDSEEEILKESQYNSDRENDE  426
```

RESULT 4
D45561
merozoite surface antigen (clone 14) - Babesia bigemina
C;Species: Babesia bigemina
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 16-Feb-1997
C;Accession: D45561
R;Mishra, V.S.; McElwain, T.F.; Dame, J.B.; Stephens, E.B.
A;Title: Isolation, sequence and differential expression of the p58 gene
Mol. Biochem. Parasitol. 53, 149-158, 1992
A;Reference number: A45561; MUID:92365724
A;Molecule type: DNA
A;Accession: D45561
A;Status: preliminary
A;Residues: 1-480 <MIS>
A;Cross-references: GB:M85187
C;Superfamily: merozoite surface antigen p58
C;Keywords: surface antigen

Query Match   34.5%;   Score 813.5;   DB 2;   Length 480;
Best Local Similarity 41.6%;   Pred. No. 1.3e-45;
Matches 180;   Conservative 74;   Mismatches 142;   Indels 37;   Gaps 12;

```
Qy      4  SDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSASV   63
```

```
RESULT 5
C45561
merozoite surface antigen (clone 9) - Babesia bigemina
C;Species: Babesia bigemina
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 16-Feb-1997
C;Accession: C45561
R;Mishra, V.S.; McElwain, T.F.; Dame, J.B.; Stephens, E.B.
Mol. Biochem. Parasitol. 53, 149-158, 1992
A;Title: Isolation, sequence and differential expression of the p58 gene family of Babes
A;Reference number: A45561; MUID:92365724
A;Accession: C45561
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-442 <MIS>
A;Cross-references: GB:M85185
C;Superfamily: merozoite surface antigen p58
C;Keywords: surface antigen

Query Match           33.5%;  Score 790;  DB 2;  Length 442;
Best Local Similarity 41.4%;  Pred. No. 4e-44;
Matches 170;  Conservative 78;  Mismatches 133;  Indels 30;  Gaps 10;

Qy    4 SDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSASV 63
         :: :::||||| :::|||::::|: :::|:|::::|  ||:||| :|:|
Db   33 AEVGDVSKTLLAANEVVNAEMEAAQINEDMKIQLANVKEITVDEVCRKDAGSPTCRKSFW 92

Qy   64 SAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEKRFW 123
         |||||||||||||||| |   ||| ||| |||||||||||:||:||:|:||:||||:|||
Db   93 IAYVDRCDEGDCLTLDSM------KYKPLSLPNPYQLDAAFMLFRESDSNPAKNEVKRFW 146

Qy  124 MRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLMATLYKTYTNVDEFGAS 183
         ||.|  .|.||:.|||.:||:|||:||.|.|.||||||:|:|:||| |||||||||::
Db  147 MRSR--SSHGDYHHFVVSLLKKNVVRDPESNDVENFASQYFYMTTLYYKTYLTVDFFAAK 204

Qy  184 FFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALP 243
         |||||:||| ||:|||:|||| ::||||::|:||  |::| ::|:::||:|:||:||| |
Db  205 FFNKLAFTTRLFGFGIQKALKRLVRSNLPVDLGT-HPEATIREIASGYEYMMTQVPAMT 263

Qy  244 KFAKRRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNNRVFIPTKKFFNKEIREPSK- 302
         .||.||||:|..|.|||||:|:|  :.|::|| :|::::|:..|| |||.||:|||::||.
Db  264 SFAGRFSKMATKTLLVTVSDVYHLPAYKRWYRKFKEEIVN-FFTDPAKLIMKHVSQPVKT 322

Qy  303 ALKEKVSTDIKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKI------G 356
         |:.: :. ..:.||:.:.|::|||::.||:..|||| :::. .:.:||:|::       .
Db  323 AYTKLVPEEHRQAIRNVVGQSTKHIAN-GVRDLSRMIKE----PSQQIIREKLPHYLSKA 377

Qy  357 QGTVDFINNEIRD--PSKALIRKVSTGCAEDLFENKIGQGTVDFINNEIRDPSKALIRKVY 414
         ||:|::.. :.    .... ::..   .:|.. |.:.|.::|.::.. ..   .|.:.:..
Db  378 KGAVEHVKVKSVVPIK---QKGDPSEAAVEETVPSG--DSAETEFEVPEEQYVDAVT 432

Qy  415 T--------EADD 419
         |        :.||
Db  433 TQEVNSEKVDADD 445
```

```
RESULT 6
C48572
rhoptry protein homolog Bc60.2 - Babesia canis
C;Species: Babesia canis
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C;Accession: C48572
R;Dalrymple, B.P.; Casu, R.E.; Peters, J.M.; Dimmock, C.M.; Gale, K.R.; Boese, R.; Wr
Mol. Biochem. Parasitol. 57, 181-192, 1993
A;Title: Characterisation of a family of multi-copy genes encoding rhoptry protein ho
A;Reference number: A48572; MUID:93165069
A;Accession: C48572
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-456 <DAL>
A;Cross-references: GB:M91168; NID:g155908; PIDN:AAA27807.1; PID:g155909
A;Note: sequence extracted from NCBI backbone (NCBIN:125083, NCBIP:125084)
C;Superfamily: merozoite surface antigen p58

Query Match           32.7%;  Score 770.5;  DB 2;  Length 456;
Best Local Similarity 35.0%;  Pred. No. 7.8e-43;
Matches 161;  Conservative 95;  Mismatches 161;  Indels 43;  Gaps 9;

Qy    1 MAPSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCS 60
            :...:..:||.:||   : .::. ...: :.|:.|.:|...:.|:.|..  ...    .
Db   31 LSKSDGAEKTLSTLLNVDASTRAALEGYRMNAAMANFSNGRREEEEAVCGNIAEETECQ 90

Qy   61 ASYSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEK 120
          ..  |.:.|...||. ||:..|||.|:.:|||||:||||:|||:.|.::| ||:|: |
Db   91 KSVAEYVESCVRVDCFSIENQKYPQEKEVQPLTLPNPYQLDAAFVFRNSESNPIKNPTE 150

Qy  121 RFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLMATLYKTYTNVDEF 180
         .|||||| |..: .|: :|.|:|.|||:  : : :|: .||.:.|:|.|.|||.: .
Db  151 AFWMRFRHGGRYGAYHNFLVNILYKNLSDSMVDDNLEGFVRKIAYMATMYYKTYTALDVV 210

Qy  181 GASEFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIP 240
         .|. :||.:|::.||. :|:|||.:||||::|  .. . .:.||.|:.. |.:|:..||.
Db  211 NARIINKIAFSRHLFGRQIRNALTNIIRSNIPEDFG-KYNVDRLRHVMGGYEEYMMKQVP 269

Qy  241 ALPKFAKRRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNNRVFIPTKKFFNKEIREP 300
          ||:||| : :.:|:. :| :. .:   .|:||:  ::.|.:.|::.|. .|:||.||:..
Db  270 SLPNFAKKYAGMVVKSLIKNVGAYQKQPWRKKLLNNQIRNFFVNKIHEPTKEFFVNKIHEP 329

Qy  301 SKALKEKVSTDIKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTV 360
             :.|.. ..:||.||:.|:||.:   |.:.     :.|.:::.:..:||.|::..
Db  330 --------TKEFFVNKIHEPPKEFFVNKIHEPT------------KEFFVNKLHEPPK 367

Qy  361 DFINNEIRDPSKALIRKVSTGCAEDLFENKIGQGTVDFINNEIRDPSKALIRKVTEADDL 420
         .||.|:.  ||:.: .. : . .:.:|.||:.  ..   .|. ...: . |:..:...:.
Db  368 EFFVNKLHEPTKEFFSNMVPGAFQKISERAGR-------HLRS--SKTVVPE--DEPSSS 416

Qy  421 FENKI---GQGTV-DFINKEIRDP-----SKALIRKVSTE 451
         :.|:.   |. .  :.. .| ..      |:.|:.:..:|
Db  417 LENEAVEDGQLTMGDVTDFEMATPTYEQGSQESLNEVGNE 456
```

```
RESULT 7
A48572
rhoptry protein homolog - Babesia ovis (fragment)
C;Species: Babesia ovis
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C;Accession: A48572
R;Dalrymple, B.P.; Casu, R.E.; Peters, J.M.; Dimmock, C.M.; Gale, K.R.; Boese, R.; Wr
Mol. Biochem. Parasitol. 57, 181-192, 1993
```

A:Title: Characterisation of a family of multi-copy genes encoding rhoptry protein homol
A:Reference number: A48572; MUID:9165069
A:Accession: A48572
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-321 <DAL>
A:Cross-references: GB:M91173; NID:g155917; PIDN:AAA27811.1; PID:g155918
A:Note: sequence extracted from NCBI backbone (NCBIN:125078, NCBIP:125081)

Query Match   26.8%; Score 632.5; DB 2;   Length 321;
Best Local Similarity 47.2%; Pred. No. 4.9e-34;
Matches 133; Conservative 45; Mismatches 87; Indels 17; Gaps 6;

```
Qy  164 LYMATLYKTYTNVDEFGASFFNKLSFTTGLFGWGIKIRALKQIIRSNLPLDIGTEHSVSR 223
            |||| :||:  |  ::| ::|  ||| |||||||||||||||||| |: |||||||||||
Db    2 LYMAATYYKTYLIVNNSDAKFFNRIAFATKIFGCIKKALKDIVRSNVPEYMG-EHSIER 60

Qy  224 LQHITSSYKDYMDTQIPALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVN 283
        |:|: ||||||||||||||||||||| |||:|||||||||||||||||:|:|||||:||||
Db   61 ISHLAHGYKDYMLTQVPTLSKFPAERYSDMVWKVLLSLAGYVKAPWYKRWFNKVKDFFVN 120

Qy  284 RVFIPTKKFFNKEIREPSKRA-LKEKVSTDTDKDLFENKIGQGTVDFFNKEIRDPSKALKE- 341
        :||||| |||||||||||||   :||||: || |  |:||||||||||||||| :
Db  121 KIGKPTKEHPHK--KHPRTAEFFDKMHERTKDFFENKIGAPTKDFFENKIGAPTKDFFEN 178

Qy  342 KVSNDAKDLFENKIGOGTGVDFINNEIRDPSKALIRKVSTGAEDLFENKIGOGTVDFINNE 401
        ||| :||  |||||||| :| |||  ::|: |||||||||||||:|:
Db  179 KIGAPTKDFFENKIGAPTKDFFENKIGAPTK------DFFENKIGAPTKDFFENK 227

Qy  402 IRDPSKALI-RKVYTEADDLFENKIGOGTVDFINKEIRDPSK 442
           ||||  |: ||||| |||||||||| :| :
Db  228 LPERIKDFFENKLPEHTKDFFENKLPEHTKDFFENKIGAPIK 269
```

RESULT 8
B48572
rhoptry protein homolog Bv60.5 - Babesia ovis (fragment)
C:Species: Babesia ovis
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 03-Nov-2000
C:Accession: B48572
R:Dalrymple, B.P.; Casu, R.E.; Peters, J.M.; Dimmock, C.M.; Gale, K.R.; Boese, R.; Wrigh
Mol. Biochem. Parasitol. 57, 181-192, 1993
A:Title: Characterisation of a family of multi-copy genes encoding rhoptry protein homol
A:Reference number: A48572; MUID:9165069
A:Accession: B48572
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-192 <DAL>
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:125078, NCBIP:125082)
C:Superfamily: merozoite surface antigen p58

Query Match   15.5%; Score 366; DB 2; Length 192;
Best Local Similarity 43.3%; Pred. No. 5.7e-19;
Matches 68; Conservative 30; Mismatches 59; Indels 0; Gaps 0;

```
Qy    1 MAPSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCS 60
Db   31 LAPAEVGNVAATLSTADEISSHDYDKELINDMRDQMREMGENFVDTVCTKAPEDSNCR 90

Qy   61 ASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEK 120
Db   91 QMVALYADRCEMYQCCLQIDNVNYPVDEEYQPLSLPNPYQLDAAFTLFKNSASNPAKNGVK 150

Qy  121 RFWMFRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIE 157
Db  151 GQWMRFNGKKHGDYHRFLIGLLAKGVTRKDGATDLE 187
```

RESULT 9
T2841
ORF MSV250 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus

C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28411
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: 220484; MUID:99102612
A:Accession: T28411
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-441 <AFO>
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97726.1; PID:g4049766
A:Genetics:
A:Note: MSV250

Query Match   6.1%; Score 143.5; DB 2; Length 441;
Best Local Similarity 21.4%; Pred. No. 0.051;
Matches 93; Conservative 58; Mismatches 142; Indels 141; Gaps 22;

```
Qy   30 INSDMSDYLSAVSDNFAERICSQVPKGSNCSASVSAYMSRCAKQDCLTLQSLKYPLEAKY 89
        |:|| ::|:   :|  ::
Db   94 INYEINDQIVPINEN----ILSYIPKEDYDKIKIINHYEFCHE------ 132

Qy   90 QPLTLPDPYQLEAAFILFKESDANPANSTEKRFWMFRPRRGKNHSYFHDLV---------- 139
                  :|:: ||  :| ::      ||:|| | ||||||||||
Db  133 ------DFYFDDYAFCEEHKFCQNHMN-YENGIWK-------ENYECYNICNCEEESKIFK 180

Qy  140 --------FNLLEKNVTRDADATD-IENFASRYLYMATLYYKTYT-NVDEFGASFFNKLSF 190
             |: ||:| : : ||  || :|:|:|
Db  181 FENKLENYTLLSKLLSSEITTTETVEN------ANTTILPFTKTDNIDDDSEYFYDREIY 233

Qy  191 TTGLFGWGIKIRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALPKFAKRFS 250
Db  234 YGYF------------TMHENKDYDFINKVVDDYVDNYV---NKFVDDFV 269

Qy  251 LMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREPSKALKEKVST 310
Db  270 NKVVNK---NVDNVVD------EVVNKNVDKV---VNEVVNKNV--DKVVDEVVNK 311

Qy  311 DTKDLFENKIGOGTVDFDFENKEIRDPSKALKEKVSNDAKDLFENKIGOGTVDFINNEIRDP 370
Db  312 NVDD-FVNKIDD---DFVNK------IDDDFVNKIDD---DFVNKNVDE- 347

Qy  371 SKALIRKVSTGAEDLFENKIGOGTVDFINNEIRDPSKALIRKVYTEADDLFENKIGOGTV 430
Db  348 --VVNKI----DDEFVNK---NVDEVVNKIDDVLNKNVNVDVNDFVNKIDDEVV 396

Qy  431 ------DFINKEIRD 439
Db  397 NKNVNDFVNKNVND 410
```

RESULT 10
B70232
hypothetical protein BBG10 - Lyme disease spirochete plasmid G/lp28-2
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: B70232
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: B70232
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-1098 <KLE>
A:Molecule type: DNA
A:Cross-references: GB:AE000786; NID:g2690008; PIDN:AAC66075.1; PID:g2690032; TIGR:BB
A:Experimental source: strain B31
A:Genetics:
A:Genome: plasmid

```
Query Match       5.7%; Score 134; DB 2; Length 1098;
Best Local Similarity 18.4%; Pred. No. 0.69;
Matches   61; Conservative   75; Mismatches 124; Indels  72; Gaps 13;

Qy  167 ATLYXYTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQH 226
             : :| :  :| |::: :|| |:|  ||| ||
Db  361 ATIYLITT------KLGELVAYLAGKWVQGMKMLFKAITTPFVKLIYEGIRXVVIGKLKE 411

Qy  227 ITSSYKDYMDTQIPALPKFAKRFSLMV--VQRLLATVAGYVDTPWYKKWYMKLKNFMV-- 282
                              |:|| |  |  || |:||
Db  412 LAT----YVSSGF-LDRFKGAFEPIIKWVQKLIDMIS-------KAYTRLKNLVTFW 456

Qy  283 --NRVFIPTKKF-FNKEIREPSKALKEKVSTDFKDLFENKIGQG-- 323
            :| |::|:| :  :||| :| |:| ||| || ||
Db  457 KKAEKEDTGSGSEPERDKKFFDPNAKTININKKMAEDYQKLQDEIFNRQRDIY-NKTGKARE 515

Qy  324 TVDFFNKEIRDPSKALIRKVSNDAKDLF------ENKIGQGTVDFINNEIRD 369
            |:| :|||  |: |:    ||:  :|        ||||| |:|
Db  516 QALRNLEKTINEKNQKPKDEYSKIFDQLTDENKKILVGVEKSVNEFNNSNYDFVNE--- 571

Qy  370 PSKALIRKVSTGAEDLFENKIGQGTVDF---INNEIRDPSKALIRKVYTEADDLFENKIG 426
            |: |:    ||: :|  ||| ||||        ||:|| |:|:
Db  572 -YQNLIKEKESREREIIKTLPHTDQVSALQKLNDEINEKNKAFVEKYGKSFETL--NESN 628

Qy  427 QGTVDFINKEIRDPSKALIRKVSTEADNLLEK 458
            || ||| :|||||| |:|:|| |||||:||
Db  629 RQVVVALEKQVNEYEKTALDRSFVEAQKALQK 660
```

RESULT 11
PC6003
surface membrane protein lmp4 - Mycoplasma hominis (fragment)
N;Alternate names: hypothetical 624 protein; Lmp4 protein
C;Species: Mycoplasma hominis
C;Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 11-Jan-2002
C;Accession: PC6003
R;Ladefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.
J. Bacteriol. 178, 2775-2784, 1996
A;Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene system
A;Reference number: JC6009; MUID:96213016
A;Accession: PC6003
A;Molecule type: DNA
A;Residues: 1-624 <LAD>
A;Cross-references: EMBL:X95601; NID:g1197335; PIDN:CAA64859.1; PID:g1197337
C;Genetics:
A;Genetics:
A;Gene: lmp4
A;Genetic code: SGC3
A;Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homology
A;Keywords: duplication; membrane protein

```
Query Match       5.6%; Score 133; DB 2; Length 624;
Best Local Similarity 22.4%; Pred. No. 0.39;
Matches  114; Conservative   85; Mismatches 171; Indels 138; Gaps 30;

Qy    2 APSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSA 61
            | :: |  :|::|            ||       |:
Db   98 AQADNLARSTEQL--NKSI-SSANTLL--AKLTD----KDNTIQQAKTELEK------ 141

Qy   62 SVSAYMSRCAKQDCLTLOSLKYPLEAKYQPLLPLPDPYQLEAAPILFKESDANPANSTEKR 121
            ::                              : :| |  :||::|:
Db  142 EVOKANQAVASNNTASMQSAKSSLDAKVTEIT----KKLE-TFNKDKFKELQQTRKN 196

Qy  122 FWMRFRRGKNHSYFHDLVFNLLEK-----NVTRDADATDIENFASRYLYMATLYKTYTN 176
            : :::|          :||  :|||     |||:|   |:
Db  197 IDEFINTNKTNPNYSELISQLTSKRDSKNSVTNSSNKSDIET-ANTELKQALA--KANTD 253

Qy  177 VDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMD 236
            |:|:   ||:|  :|
Db  254 KDQ-----ADNLA------KSAKEQL 287

Qy  237 TOIPALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKF--FN 294
            | :  :| :|     :|:||:|  |:
Db  288 TIQQAKTELEK-----------VQKANQAVASNNTAS-----MQSAKSSLDAKVTEITKKLETFN 337
```

```
Qy  295 ------KEIREPSKALKE---KVSTD-----------TKDLFENKIGQGTVDFDFN--KEI 332
            :: |  | |:|      |::|           |:| ||| ||| |::  : : :
Db  338 KDKDAKFKELEQTRKDIDEFIKQIENDPQTKKNYQNIVKNLKDKKRAEKNSITFSNNKKEI 397

Qy  333 RDPSKALKEKVSNDAKDLFENKI-GQGTVDFINNEIRDPSKALIRKVSTGAEDLFE--- 387
            :| |:  :|  ::  |   : |:  :  ||: :|   :| | |:||
Db  398 QDANKSLQDELNN------AKITRKGITDFYNSK----KQL----EDLIKTDDA 437

Qy  388 NKIGQGTVDFI------------NNEIRDPSKAL--IRKYY-TEADDLFENKIGQ----- 427
            ::  :  :|:              :|||| |: |  |:: : |||:|:|:|||
Db  438 KKVGTTEADTILDHYKNISDASKNEEIKQATQKINDIKIIETKIQEKKRNEFSQFEQIK 497

Qy  428 -GTVDFINKEIRDPSKALIRKVSTEADN 454
            |:|:| :|:|| |: |:|:|| |||||
Db  498 NELQSFINDLKDQKYNSIR---TKIEN 522
```

RESULT 12
S48326
hypothetical protein YML071c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 02-Dec-1994 #sequence_revision 09-Mar-1996 #text_change 29-Oct-1999
A;Accession: S48326; S48825
R;Bowman, S.; Churcher, C.
submitted to the EMBL Data Library, August 1994
A;Reference number: S48326
A;Molecule type: DNA
A;Residues: 1-113 <BOW>
A;Cross-references: EMBL:Z38114; NID:g558402; PIDN:CAA86249.1; PID:g558603; MIPS:YML0
R;Brown, D.; Bowman, S.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48816
A;Accession: S48825
A;Molecule type: DNA
A;Residues: 88-607 <BRO>
A;Cross-references: EMBL:Z46373; NID:g587529; PIDN:CAA86507.1; PID:g587539; MIPS:YML0
C;Genetics:
A;Map position: 13L

```
Query Match       5.4%; Score 127; DB 2; Length 607;
Best Local Similarity 22.1%; Pred. No. 0.92;
Matches   90; Conservative   63; Mismatches 147; Indels 108; Gaps 21;

Qy  108 KESDANPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDI---------EN 158
            ::   ||| |:|      :||::||:|::|    :| |: |::
Db  137 EDNDTGRIMTTESN---NLARKKEDEFHKALSRLRNRISTKEDDKDDIRSDTLVTVLEN 193

Qy  159 FAS---------------RYLYMATLYYKT--------YTNVDEFGASFFNKLSFT 191
            |::               :||:|||||:|        |||||:|:|:: :|
Db  194 LDSITDLMELPFLARTCIRTGHYQEAVMLYTHTTSLRSRPGCSTIVDEVCEKVLNEIS-T 252

Qy  192 -GIKRALKQIIRSNLPLDIGTEHSV-----SRLQHITSSYKDY 234
            ||::|||::|:|| |:||:|||||      |:|:||  ||:
Db  253 TMLSCLVKLLSTNVSVNSLKKIL-QYLNSIPPFDGKTNKSLLSVFLAMRXFITDEIASY 311

Qy  235 -MDTQIPALPKFAKRFSL--MVVQRLLATVAGYVDTPWYKKWYMKLKNFM--VNRVFIPT 289
            |  :|  :|:|  |:||  :||||:||| |:| ||| |:| |: |::|:  ||||::|
Db  312 PLDVE-------SSNESLIEMVVRRKIEVLREHV----YMSLNVFLKSFLYDTNDLEIP- 359

Qy  290 KKFFNKEIREPSKALKEKVSTDTFKDLFENKIGQGTVDFINKEIRDPSKALKEKVSNDAKD 349
            ::|:|||:||||| |:| ||||: |:|||||| ||:|||::||:|||||| |:||::
Db  360 ---FPEELE--STVLRINGTNEEKIEIEEKE------KETKEEEYOKQDSVANEED 404

Qy  350 LFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFINNEIRDPSKAL 409
            :||||| ||| ||::|:|  |                :||||||:|||::||:||||:|
Db  405 -VTENR-----SIEDVQEEV----QGKVEGEDDGAERTENEIENEIV----NKTEDKAE-- 450

Qy  410 IRKVYTEADDLFENKIGQGTVDFINKEI--RDPSKALIRKVSTEADNL 455
            |:||:|||:|:|::|||     |::::|: :|||| :|:|||||:|:
Db  451 --KEKEEEVNTKDNKAEKEEEEEINKVEVTPEEPSKSIDNKAEKEEEEI 496
```

RESULT 13
E82909
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1830 <GLA>
A;Cross-references: GB:AE002127; GB:AF222894; NID:g6899268; PIDN:AAF30701.1; GSPDB:GN001
A;Experimental source: serovar 3; biovar 1
C;Genetics:
A;Gene: UU292
A;Genetic code: SGC3

Query Match   5.8%; Score 125; DB 2; Length 1830;
Best Local Similarity 22.9%; Pred. No. 5.1;
Matches 107; Conservative 68; Mismatches 195; Indels 98; Gaps 22;

```
Qy     17  ASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSASVSAVMSRCAKQDCL   76
               :|| |  |  | |||  |  |  | |  ::| |:| | | |    |:|  | | :|
Db    401  AQQKTASIANAIYDAWD-SEYFNKDSDSFRQQI---------------GFPANNTNQDLV  443

Qy     77  --TLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEKRFWMRFRRGKNHSY  134
             |:  |    | |  ||| |   |||||||||||||||||||||||||||||| | |
Db    444  EKTLGSSKG--AAFWAKTALENDEELRKSPSNFKNT--------STIVKDWLISNKRRFGSF  495

Qy    135  FHDLVFNL----LEKN--VTRDADATDIENFASRYLYMATLIYKTYTNVDEFGASFFNKL  188
            |||||||     |||   | ||||||||||||||||||||||||||||||||| |:|
Db    496  VNLITSNLITSIGLDKNGYISQFLSRKDADNTNNLYKLASALASTPY--IDMVISGIONDL  553

Qy    189  S----FTTGLEGWG--IKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPAL  242
            |     | |  |     ||| ||  |||  ||||  ||  |  | | |:| || | ||
Db    554  DNLDLDKNNIIGTQNISKKILYDVFRKNIINDQNLYELIKGMFH--SQLFNENKSQIKSL  611

Qy    243  PKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKN---EMVNRNVFIPTKKFFNKEIRE  299
            |||  |||||||||||||||||||||||||||||:    |  || ||| |  || |:|
Db    612  IK-ALLSDLVSSHKLLEIIFGEFATD---------------NLKNNNEFLKNLLIIQNK----  656

Qy    300  PSKALKEKVSTDTKDLFENK--IGQGTVDFFNKEIRDPSKA-LKEKVSNDAKDLFENK-  354
            |:|:    |:| ||||||||   || | |  |  |||  |  ||||||| ||  |||||
Db    657  QTKSLDNIINVFVDALFDNDNEYLLQGPLSNVLAKIVKTNKALLTTRIDNLFNDLFKDQT  716

Qy    355  -------IGQGTVDFINNEIRDPSKALIRK--VSTGAEDLFENKIGQGTVDFINN  400
                      || | ||  |||  |||   | |  |   |||||  || | |||||
Db    717  TLDATYNILIRVLGQNIPDF----IKTPQASLIKKILSKGTNLKLINTLKINTIDLLTD  772

Qy    401  EIRDPSKALIRKVVTEADLLFENKIGQGTVDFINKEIRDPSKALIRKV  448
            ||||| |||  |:    ||||||||||||||||||||||||||||||
Db    773  S--------------KTYAGIFDSSETSI----INLIQKSIDFTSKSFISDV  806
```

RESULT 14
E70132
hypothetical protein BB0261 - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: E70132
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.J. Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: E70132
A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA
A;Residues: 1-460 <KLE>
A;Cross-references: GB:AE001137; GB:AE000783; NID:g2688160; PIDN:AAC66691.1; PID:g268
A;Superfamily: hypothetical protein BB0261; tetratricopeptide repeat homology
C;Experimental source: strain B31
F;103-136/Domain: tetratricopeptide repeat homology <TT1>
F;137-170/Domain: tetratricopeptide repeat homology <TT2>
F;205-238/Domain: tetratricopeptide repeat homology <TT3>
F;239-272/Domain: tetratricopeptide repeat homology <TT4>
F;278-310/Domain: tetratricopeptide repeat homology #status atypical <TT5>
F;311-344/Domain: tetratricopeptide repeat homology <TT6>

Query Match   5.8%; Score 122.5; DB 1; Length 460;
Best Local Similarity 23.5%; Pred. No. 1.3;
Matches 93; Conservative 55; Mismatches 144; Indels 103; Gaps 20;

```
Qy    113  NPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLIYK  172
            ||||||||||||||||||||||||| || |||||| |||  |||||||||||||||| |
Db     94  NPINQEINTF---------------KAFLSYGFYNLKRNFNRE------ALEYLRKAFILKK  134

Qy    173  TYTNVDEF-GASFFNKLSFTTGLFGWGIKRALKQI----IRSNLPLDIGTEHSVS  222
            || ||||  ||||||||||||||||||||||| |     ||||||||||||  |
Db    135  TDEDANYYLGVAFLIKNEMYKEALY-----YLKKVHNENKNKRDVLKHLGITLFNLESYR  188

Qy    223  RLQHITSSYKDYMDTQIPALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMV  282
            | |:|||||| |||  | |  ||  ||||||||||||||||| || |||| | ||||
Db    189  QAVIIFNNIKNIIQGDIEALLAYAKSLSKMNQDHLALEIANKI---------KQKDGMI  238

Qy    283  NRVFIPTKFF--NKEIREPSKALKE--KVSDTK-------DLFENKIGQ--  323
            |||| |  |   |||| | |||| |  |||                   | ::|||| |
Db    239  YEALLITSEIHSKNKELEKLEQNIKEIIKVKPDLPKKTFLELFYN-LGELQISVENYQKA  297

Qy    324  TVDFFNKEIRDPS-KALKEKVSNDAKDLFEN-------KIGQGTVDFINNEIRDPSKALI  375
            ||||  | | |   || |||  |  ||| |        ||  | ||||||  |||||||
Db    298  TEAFTKVEEIIDPNYKNIKEKL-EFSKRLNENIALRIYLRSSKENFEKIANEI----  349

Qy    376  RKVSTGAEDLFENKIGQGTVDFI--------NNEIRDPSKALIRKVYTEAD--DLFENKI  425
            ||||  |  |||||  || ||| |         |||||||||||||| |   ||||||
Db    350  LKLYLNKFQIRDSKINEVTTQFIDINFHLANNQWEE--NLIVRFVRTEQDTFGELF----  403

Qy    426  GQGTVDFINKEIRDPSKALI----RKVSTEADNLLE  457
            |:|||||||||||||||| |    ||   | |:| |
Db    404  --LKDFISKIKENKIKGLCIAPSKFSLKAKQMIE  435
```

RESULT 15
S35027
cytotoxin RTX homolog frpC - Neisseria meningitidis
C;Species: Neisseria meningitidis
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 08-Oct-1999
C;Accession: S35027
R;Thompson, S.A.; Wang, L.L.; Sparling, P.F.
Mol. Microbiol. 9, 85-96, 1993
A;Title: Cloning and nucleotide sequence of frpC, a second gene from Neisseria mening
A;Reference number: S35026; MUID:94018616
A;Accession: S35027
C;Genetics:
A;Gene: frpC
A;Molecule type: DNA
A;Residues: 1-1829 <THO>
A;Cross-references: GB:L06299; NID:g293961; PIDN:AAA99902.1; PID:g293963

Query Match   5.2%; Score 122; DB 2; Length 1829;
Best Local Similarity 20.3%; Pred. No. 8;
Matches 80; Conservative 63; Mismatches 156; Indels 96; Gaps 18;

```
Qy     94  LPDPYQLEAAFILFKESDANPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDA--  151
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     35  LGDRYADNAAAIVGKDTNLNGLN-----LWM--KKG--------VENLWDTVGKKTRL   78

Qy    152  ---DATDIENFASRYLYMATLIYKTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIR  208
                |:|||||||||||||||||||||||||||||||||||||||||||||||::|
Db     79  EKFDRVALQHF-SQYVDLINKNNGRLPNTSEIERSYKKAVTY----HGVSSSAAIDLVIN  133
```

```
Qy  209 SNLP--------LDIGTE---HSVSRLQHITSSVYKDYMDTQIPALPK------FAKRFS 250
            |||              || :||                           ||||:
Db  134 RSLPDMADGYWALGLGIEAERIHNEQAVNNPNGSERDNRKQLISALDKGPFDGSFKEKHFT 193

Qy  251 LM--VVQRLLATVAGYVDTPFWYK--KW----YMKLKNFMVNRVFIPTKKFFENKEIREPSK 302
        |:    |    |  | | ||  ||  |      |||| |  |   |      ||||| | :
Db  194 FLQSVMMDLTKLGVEYTIDGWQKIGGWGNGIINDLYKSVVKREWTGIFEIVNNNIKQGNE 253

Qy  303 ALKEKVS--------------TDTKDLFENKIGQGTVDFENKEIRDPSKA 338
         :| ::                 | ||||||||||||||||| ||:||||
Db  254 AFKNEINSLVHDMKAAGKEFGDDLNTQWNNLTQAAEIIYNDIVDNTSQGIEKGVK-AIKE 312

Qy  339 LKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVYSTGAEDLFENKIGQGTVDFI 398
        | :|| : | ||| :                  | :   | |||  |    |
Db  313 LSEKMKNAASDLADGS-------AEKAKQVVEDLAQAAKEAYEN-AKSTAEKA 357

Qy  399 NNEIRDPSKAL--IRKVYTEADDLFENKIGQGTVD 431
         | : :   |    :  |    ||| ||
Db  358 AQAAREFFKGLPSKFDLAEKFRDLFPNP--EGWID 390
```

RESULT 16

HMIVUR

hemagglutinin precursor - influenza A virus (strain A/USSR/90/77)

```
C;Species: Influenza A virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 16-Jul-1999
C;Accession: A04064
R;Concannon, P.; Cummings, I.W.; Salser, W.A.
J;Virol. 49, 276-278, 1984
A;Title: Nucleotide sequence of the influenza virus A/USSR/90/77 hemagglutinin gene.
A;Reference number: A04064; MUID:84090410
A;Accession: A04064
A;Molecule type: genomic RNA
A;Residues: 1-566 <CON>
A;Cross-references: GB:K01330; NID:g324193; PIDN:AAA43206.1; PID:g554652
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-343/Product: hemagglutinin HA1 chain #status predicted <HA1>
F;345-566/Product: hemagglutinin HA2 chain #status predicted <HA2>
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted
```

```
Query Match          5.1%;  Score 121;  DB 1;  Length 566;
Best Local Similarity  20.8%;  Pred. No. 2.1;
Matches   94;  Conservative   68;  Mismatches  161;  Indels  130;  Gaps  23;
```

```
Qy   33 DMSDYLSAVS--DNF---------AERICSQVPKGSNCSASVSAVMSRCAKQD-----CLTLQS 80
         : |  ::|   ||         ||||   |       |     | ||  :           | :|
Db  117 ELREQLSSVSSFEREIFPKER--SWPK-HNVTRGVTASCSHGKRSSFYRNILWLTEKN 172

Qy   81 LKYPLEAKYQPLTLPDPYQLEAAFILFKESDA------NPANSTEKRFWMRFRRGKNHS 133
         | ||           ||        || :|        | |:: | |  | | |
Db  173 GSYP---------NLSSSYVNNKEKEVLVLWGVHHPSNIEDQKTIYR---KENA 214

Qy  134 YFHDLVFNLLEKNVTRDADATDIENFASRLYLMATLYKYHYTNVDEFGASF---FNKLSF 190
          |  | :    : |          :  || |  :   | :|  ||    |      | |:|
Db  215 YVSVVSSNYNRRFTPEIAERPKVRGGAGRINYYWTLLEPGDTIIFANGNLIAPWHAFAL 274

Qy  191 TTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSS--YKDYMDTQIPALPKFAKR 248
         | |   |         |  ||          :     | |     : | |  | |
Db  275 IITSNASMDECDTKCQTPQGAINSSLPFQNIHPVIIGECPKYVRS 326

Qy  249 FSLMVV------------QRLLATVAGYVDTPFW---YKKWY-----M 275
             :|             ||| |  ||  | |     || |     |
Db  327 TKLRMVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMIDGWYGYHHQNEQGSYAADQKSTQ 386

Qy  276 KLKNPMVNRV------------FIPTKKFFNEIREPSKALKEKVSFDTKD-------L 315
          | |  |:|:             ||   | |  |     |  ::    |        |
Db  387 NAINGITNKVNSVIEKMNTQPTAVGKEFNK-LEKRMENLNKKVDDGFLDIWTYNAELLVL 445

Qy  316 FENKIGQGTVDFENKEIRDPSKALKEKVSNDAKDLFENKIGQGTVDF---INNEIRDPSK 372
          |:    |  ||      ||   |  | :: |:|  |      |         | |
Db  446 LEN---ERTLDFHDSSVKNAYEKVKSQLKNNAKE------IGNGCFEFYHKCNNE----- 491

Qy  373 ALIRKVSTGAEDL----FENKIGQGTVDFINNE 401
         |       |  |     |:    | |
Db  492 -CMESVKNGTDYDYPKYSEESKLNREKIDGVKLE 523
```

RESULT 17

D71623

erythrocyte membrane protein PfEMP3 PFB0095c - malaria parasite (Plasmodium falciparu...)

```
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: D71623
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743
A;Accession: D71623
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2441 <GAR>
A;Cross-references: GB:AE001371; NID:g3845092; PIDN:AAC71809.1; PID:g384
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0095c
```

```
Query Match          5.0%;  Score 118.5;  DB 2;  Length 2441;
Best Local Similarity  31.4%;  Pred. No. 20;
Matches   58;  Conservative   32;  Mismatches  64;  Indels  31;  Gaps  14;
```

```
Qy  294 NKEIR-EPSKALKEKVSFDTKDLFENKIGQG---TVDFENKEIRDP-SKALKEKVSNDAK 348
        |||:  |||  |||| |       |  |  |   || |   | |   |||||| |
Db 1121 NKELQNKGSEGLKENAELKNEL-QNKGSSEGLKENAELKNKELRNKGSEGLKENAELKNK 1179

Qy  349 DLFENKIGQGTVDFINNEIRDPSKALIRKVYSTGAEDLFE------NKIGQG---TVDFI 398
         : :  ||    |        | |||  :|  ::  | :|       ||| |     || |
Db 1180 EL-QNKGSEGLKE--NAELK--NKELQNKGSEGLKENAELKNEL-QNKGSEGLKENAELK 1234

Qy  399 NNEIRDP-SKALIRKVYTEADDLFENKIGQGTVDFINKEIRDPSKALIRKVYSTEAD---- 453
         |:: |   |||  : :     :|:   ||   | |   | |   :  |:  |:: ||
Db 1235 NKELRNKGSEGLKENVYIN-NDLKNNDI-QNKDLSNKDMK--NKELLNKDISNKDMKNK 1289

Qy  454 NLLEK 458
         |:|:
Db 1290 ELLNK 1294
```

RESULT 18

A72358

conserved hypothetical protein - Thermotoga maritima (strain MSB8)

```
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: A72358
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A;Reference number: A72200; MUID:99287316
A;Accession: A72358
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-447 <ARN>
A;Cross-references: GB:AE001734; GB:AE000512; NID:g4981107; PIDN:AAD35685.1; PID:g498
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0600
```

```
Query Match          5.0%;  Score 117.5;  DB 2;  Length 447;
Best Local Similarity  21.8%;  Pred. No. 2.6;
```

Matches 74; Conservative 50; Mismatches 104; Indels 111; Gaps 20;

```
Qy    120 KRFWMR-FRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFAS-------RYLYMATLY----- 170
            | ||   |||||||||||||                    |||          ||||||
Db    122 KRFYHHGFEPVKSY-FFH------LGTDILKETNKNIVEYFSSPENIEFLFGTTVVEDVEF 174

Qy    171 -YKTYTNVDEFGASFFNKLSFTTGLFGWGIKRALK------QIIRSNLPLDIGTEHSVSR 223
            |||| ||||||| |||| | :  :|: :|:         ||||||||||| || |
Db    175 GPPHRVHINRGDFE---FDELVIAVGRSGHRLMERLKEKYPQLVRPNQFVDIGVRYELP- 230

Qy    224 LQHTTSSYKDYMDYQI------------------PA----LPKFAKRFSLMVYQRLLATVAG 263
            | :| :  :              ::  |:    ||        ::     ||
Db    231 -NHVMDPFSDMYEVKVRYRTKTGYICRIFCQNPAGKVTLEKYED-----------FTTVNG 279

Qy    264 YVDTPWYKKWYMKLKNFMWVNRVFIPTKFFENKEIREPS----------KALKEKVST 310
            |  :   ::  :|   |   ::| :  :| |||            || ||| :|
Db    280 YSDS-----LHKTEN--TNFAILVTTR-FTEPEFKDPTGYGVNLAKLANILAGDKEKVIL 330

Qy    311 DTKDLFENKIGQGTVDFENKEIR------DPSKALKEKVSNDAKDLFENKIGQGTVDF 362
            :  ||||||||||||||  :       |||||||||:|::|   ||||||||||||
Db    331 QTYG------------DF-KEFRRTKRLGRVHPTLDSESYILGDANLVFPSKIRESLVDF 377

Qy    363 INNEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFINNE 401
            :   :  | :| :  :     ||  :  : || :| :
Db    378 VEN--------LEKVIPGA-SYFDNLIYAVEVKFYTNK 406
```

RESULT 19
H90565
restriction modification enzyme subunit r2 [imported] - Mycoplasma pulmonis (strain UAB
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Accession: H90565
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-986 <KUR>
A;Cross-references: GB:AL445566; PID:gl4089846; PIDN:CAC13605.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MPU_4320
A;Genetic code: SGC3

Query Match   5.0%; Score 117; DB 2; Length 986;
Best Local Similarity 23.1%; Pred. No. 7.7;
Matches 81; Conservative 67; Mismatches 142; Indels 60; Gaps 17;

```
Qy    140 FNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEFGASFFN--KLSFTTGLFGW 197
            |:|:::                |:          | ||::|         ||||| |||
Db    648 FSLEEQSIN---DAFKIYANSSDKEIQQLVYGEKYEQVVEDFINFWNSLKISFS------ 698

Qy    198 GIKRALKQIIRSNLPLDIGTEH-----SVSRLQHTTSSYK---------DYMDTQIPALP 243
             || : :  :        :          :  :   : |           ||  |  |
Db    699 NIYDEKNNEIFRNISLENKKKYLKNLSQVSNIFPSSLKTFEKYGNEKISDFSLEQLNQYQ 758

Qy    244 KFA---KRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMWVNRVFIPTKKFFNKEI-R 298
            |:|    :  |    |:       |  |  : :   :| :  | :| :  :| |||  |
Db    759 KYANEIKKNLSTNEKEKISYEVLNSIDISNIKFAY------KEMIIDEIYLENLLFNKKISK 815

Qy    299 EPSKAL--KEKVSVSDTK--DLFENKIGQGTVDFENKEIRDPSKALKEKVSNDAKDLFENK 354
             |||||  ||||| :    ||||||||||||||  ::  |||||||:|:::    |||||
Db    816 YPNNRLTYEDTLSEIDKHIQLIKNNYNQKI---NQKEYEIFLLVQKKNEIKNFFIKK 872

Qy    355 IGQGTVDFINNEIRDPSKALIR-----KVSTGAEDLFENKIGQGTVDFINNEIRDPSKALI 410
            ||||||||   :  ||:|| |       :   || ||||||||||||   :  ||||||
Db    873 --DKSLD--EKEFTDYGKRILKSVFQKVKVKNQIEAMWLEKILKEYHGINNDQIRKDWKKRI 928

Qy    411 RKVYTEADDLFENKIGQGTVDFINK---EIRDPSKALIRKVSTEADNLLE 457
            ||: ||::|||||||||||||||     :  ||||| :||| ||::|:
Db    929 ND--KDLDDIEKS---------EFIKKWSRSRSKEVDKDIIDKLSIEYKESIE 969
```

RESULT 20
B42641
kinesin-related protein CIN8 - yeast (Saccharomyces cerevisiae)
C;Alternate names: protein YEL061c
C;Species: Saccharomyces cerevisiae
C;Date: 03-Mar-1994 #sequence_revision 19-May-1995 #text_change 23-Mar-2001
A;Accession: S50528; B42641; S44939; S69009; S20134
R;Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
C;Description: The sequence of S. cerevisiae cosmids 9669, 8334, 8199, and lambda clo
A;Reference number: S50434
A;Molecule type: DNA
A;Residues: 1-1038 <DIE>
A;Cross-references: EMBL:U18795; NID:g603241; PIDN:AAB65026.1; PID:g603257; GSPDB:GN0
R;Hoyt, M.A.; He, L.; Loo, K.K.; Saunders, W.S.
J. Cell Biol. 118, 109-120, 1992
A;Title: Two Saccharomyces cerevisiae kinesin-related gene products required for mito
A;Reference number: A42641; MUID:92317149
A;Accession: B42641
A;Molecule type: DNA
A;Residues: 1-253,'A',255-830,'H',832-1038 <HOY>
A;Cross-references: EMBL:Z11859; NID:g93541; PIDN:CAA77885.1; PID:g93542
(NCBIN:107723, NCBIP:107726)
R;Rousselet, G.; Simon, M.; Ripoche, P.; Buhler, J.M.
submitted to the EMBL Data Library, May 1994
A;Note: sequence extracted from NCBI backbone (NCBIN:107723, NCBIP:107726)
A;Description: A second nitrogen permease regulator in Saccharomyces cerevisiae.
A;Accession: S44938
A;Molecule type: DNA
A;Residues: 1021-1038 <ROU>
A;Cross-references: EMBL:X79105; NID:g485969; PIDN:CAA55722.1; PID:g485971
R;Rousselet, G.; Simon, M.; Ripoche, P.; Buhler, J.M.
FEBS Lett. 359, 215-219, 1995
A;Title: A second nitrogen permease regulator in Saccharomyces cerevisiae.
A;Reference number: S69008; MUID:95172238
A;Accession: S69009
A;Molecule type: DNA
A;Residues: 1021-1038 <ROW>
A;Cross-references: EMBL:X79105; NID:g485969; PIDN:CAA55722.1; PID:g485971
A;Status: translation not shown
C;Genetics:
A;Gene: SGD:CIN8; MIPS:YEL061c
A;Cross-references: SGD:S0000787; MIPS:YEL061c
A;Map position: 5L
A;Superfamily: kinesin-related protein CIN8; kinesin motor domain homology
C;Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loo
F;75-520/Domain: kinesin motor domain homology #status atypical <KMOT>
F;166-173/Region: nucleotide-binding motif A (p-loop)
F;172/binding site: ATP (Lys) #status predicted

Query Match   5.0%; Score 117; DB 1; Length 1038;
Best Local Similarity 19.5%; Pred. No. 8.2;
Matches 119; Conservative 97; Mismatches 214; Indels 180; Gaps 28;

```
Qy      4 SDSVGDVTKTLIAASESVDSAANAYMI--------NSDMSDYLSAVSD------- NFA 46
           |||.|:   | : :| :|| :   :          ||  ||||:  ||
Db    188 SDAAGIIPRVLLKLFDLELQQNDYVWKCSFIELYNEELKDLLDSNSNGSSNTGFDGQFM 247

Qy     47 ERICSQVPKGSNCSASAYSAYMSRCAKODC------LTLQS--LKYPLEAKYQPLTLPDP 97
           ::|    |:: : ||:| ::|:| :          |||:   ||  ||: | || |||
Db    248 KKLRIFDSSTANTTSNSASSASSRSNSNSNSSPRSLNDLTPKAALLRKRLRTKSLPNTIKQQ 307

Qy     98 YQLEAAFILFKESDANPANSTE-------------KRFWMRFRRGKNHSYFHDLV------ 139
           ||   |  :::|| :| ::|                |||:|||||||||||||
Db    308 YQQQQAVNSRNNSSNSGSSTTNNASSNTNTNNGQRSSMAPNDQTNCIYIQNLQEFHITNA 367

Qy    140 ---FNLLEKNVT-RDADATDIENFASR--YLYMATLYYKTY------------------TN 176
              |||::|    | |       |||    ||:|||:   |                  ||
Db    368 MEGLNLLQKGLKHRQVASTKMNDFSSRSHTIFTIILYKHQDELFRISKMNLVDLAGSEN 427
```

```
Qy   177 VDEFG------------------ASFFNKLSFTTG-----------LFGWGIKRA 202
            ::| |                :  :                    ||  ||||
Db   428 INRSGALNQRAKEAGSINQSLLTLGRVINALVDKSGHIPFRRESKLTFLLQDSLGGNTKTA 487

Qy   203 LKQIIRSNLPLDIGTEHSVSRLQHTTSSYKDYMDTQIPALPKFARFSLM------V 253
            |::|      |  |   |     |  | :    | |   :|  :      |
Db   488 L---IATISPAKVTSEETCSTLEY-ASKAKNIKNK--POLGSFIMKDILVKNITMELAKI 541

Qy   254 VQRLLATVAG--YVDTPWYKWYMKLKNFMVNRVFIPTKKFFNKEIREPSKALKEKVST 310
             :|     |     ::      :|  |    |:    :        |  : |
Db   542 KSDLLSTKSKEGIYMSQQDHYKNLNSDLESYK-NEV------QECKREIESLTSKNAL 591

Qy   311 DTKDLFENKIGQGTVDFFNKEIR----------DPSKALKEKVSNDAKDL--FENKIGQGT 359
         :||:|::|        ||::|            ||:|  :     :  :   |:||||||
Db   592 LVKDKLKSK---ETIQSONCQIESLKTTIDHLRAQDLKQHKTEIEISDFNNKLQKLTEVM 648

Qy   360 ------------VDF------INNEIRDPSKALIRKVSTGAED--LFENKIGQGTVDFIN 399
                        ||:      ::|::   |    ||  | :  |||||||||
Db   649 QMALIHYKKRELDLNQKFEMHITKEIKKLKSTLFLQLNTMQOESILQETNI-QPNLDMIK 707

Qy   400 NEIRDPSKAL----------IRKVYTEADDLFENKIGQGTVDFINKEIRDPSKAL---I 445
         ||:|  |             :|::  :   |::| :         :|::|
Db   708 NEVLLMRTMQEKAELMYKDCVKKILNESPKFFNVVIEK--IDIIRVDFQKFYKNIAENL 765

Qy   446 RKVSTEADNL 455
             ::  |:|
Db   766 SDISEENNNM 775
```

RESULT 21

E81086

iron-regulated protein FrpC NMB1415 [imported] - Neisseria meningitidis (strain MC58 ser

C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: E81086
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81086
A;Accession: E81086
A;Molecule type: DNA
A;Status: preliminary
A;Cross-references: GB;AE002490; GB;AE002098; NID:g7226651; PIDN:AAF41776.1; PID:g72266
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1415

Query Match 5.0%; Score 117; DB 2; Length 1829;
Best Local Similarity 20.2%;
Matches 81; Conservative 62; Mismatches 150; Indels 108; Gaps 19;
A;Molecule type: DNA
A;Residues: 1-1829 <TET>

```
Qy    94 LPDPYQLEAAFILFKESDANPANSTEKRFWMRFRRGRKNHSYPHDLVPNLLEKNVTRIDADA 153
            | |  ||  :||||||||||||||||||  :   :|  |      |: |   |::
Db    35 LGDRYADDAAIVGKDANLNGLN------LWM---KKG------VENLWDDTVGK---K 75

Qy   154 TDIENFASRYLIYMATLYYKTYP------NVDEFGASPFNKLSFTTGLFGWGIKRA 202
            :|||||||||||||||          || : |:||||| | || ||||||
Db    76 TRLEKFDR------VALQHFRQYARLINQNGRLPNTSEIERSYYKAVTDN---GVSSSAA 127

Qy   203 LKQIIRSNLP----------LDIGTE---HSVSRLQHTTSSYKDYMDTQIPALPK---- 244
         |||||||:||          ||||      :| ||  ::||  ||  ||:||:|
Db   128 IDLVIRSLPDMADGTWALGIGIEAERIHNEQAVNPNGSERDNRKQLISALDKGFDGSF 187

Qy   245 FAKRFSLMVVQRLLATVAG------YVDTPWYK---KW----YMKLKNFMVNRVFIPTKKFFNKE 296
         |||  :|||||::||||        ||||  ||   ||    ::||::| ::||::|||   | |
Db   188 KEKHFFLQSVMMDVTKLGVEYTIDGWQKIGGWGNGIINDLYKSVKREWTGIFEIVNN 247

Qy   297 IREPSKALKEKVS------
         |::    |::|
Db   248 IKQGNEAFKNEINSLVHDMKAAGKREFGDDLNTQWNNLTQAAEIIYNDIVDNTSQGIEKGV 307
```

```
Db   333 RDPSKALKEKVSNDAKDLFENKIGQGTVDFINEIRDPSKALIRKVSTGAEDLFENKIGQ 392
Qy   308 K-AIKELSEKMKNAASDLADGS-----------------AEKAQVVEDLAQAAKEAYEN--AK 351
         :  ::|   :|: ::||  | |                  :::  || || :::: : |
Db   393 GTVDFINEIRDPSKAL---IRKVYTEADDLFENKIGQGTVD 431

Db   352 STAEKAAQAAREFFKGLPSFPKDLAEKFRDLFPNP--EGWID 390
Qy   393 GTVDFINEIRDPSKAL---IRKVYTEADDLFENKIGQGTVD 431
```

RESULT 22

G70163

hypothetical protein BB0512 - Lyme disease spirochete

C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C;Accession: G70163
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: G70163
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2166 <KLE>
A;Cross-references: GB;AE001153; GB;AE000783; NID:g2688419; PIDN:AAC66876.1; PID:g268
A;Experimental source: strain B31

Query Match 5.0%; Score 117; DB 2; Length 2166;
Best Local Similarity 19.3%;
Matches 105; Conservative 94; Mismatches 182; Indels 164; Gaps 24;

```
Qy    16 AASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSASVSAYMSRCAKQDC 75
          | ||:|||:::   :::|:||||:|:|||:|: |:                |: ||:|
Db   500 ALFESLDSDSSSKFE-NQMESKYKS----FTDKLTFAGMDEFSLMYGEKFETLSQEATNNY 553

Qy    76 LTLQSIKPPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEKRFWMRFRR-------- 128
           :|:|:|         | | |  |  : |:::||::|:: :|  :|:
Db   554 QEFQDLNKKLE---------NEIESFYNMFEK-------TQETLKVDFNTSLINIKD 594

Qy   129 --GKNHSYFHDLVFNLLEKNVTR--------DADATDIENFASRYLMATLYY 171
           ||::    : :|: :|::::          |||  ::||:   ::|
Db   595 EIGKNIVEERDRYDEVNIFVTQLEESKLQYSKWQGEMDSNLKNIESQINK------ 645

Qy   172 KTYTNVD-----------EFGASFFNKLS----FTTGLFG-WGIKRALKQIIRSNLP 212
           |: |:           |:   |::|:     |  |:|  ||:::|::|
Db   646 ---TNEEFLSLIQIQKDKGIELESESVFNDLSDHIQKKAIDMHGSW--KDELIALNKSLLD 700

Qy   213 LDIGTEHSVSRLQ-HITSSYKDYMDTQIPALPKFAKRFSLMVVQRLLATVAGYVDTPWYK 271
          :|: |     :| :|  | :|   : | | :|   :   :    ||   ::|   ||
Db   701 IKVSSEELLSSATLKIESLEKDVNDRMEYVLKTGDIESLVIEK------ 744

Qy   272 KWYMKLKNFMVNR--VFIPTKKFFNKE---IREPSKALKE--KVSTDFKDLFENKIGQ 322
          :|:|::| :   :  | :: ::| ::    :::|||:|:    |:|:|:||||||||
Db   745 --YKELKDMSYSQSDEAILGIKEFINRQTEIIKDKSVFMLEDLNKKFDDKNNFVISKIEE 802

Qy   323 GTVDFFNKEIRDPSKALKEKVSNDAKDLFENKI---------GQGTV 360
           |:: | |:|  | ::|::::|::| ||:|
Db   803 --CDYKLKDFKIESEDILNNFKSDLNEFIESKLQIVSNIKSDNQKIDDFDLRISKDILN 860

Qy   361 --DFINEIRDPS-----KALIRKVSTGAEDLFENKIGQGTVDFINEIRDPSKAL----- 409
            |: :: :| |     ::|:|::   |:||:|::||   |||:|::|
Db   861 RKDSINEVDSKLSDWQSKLNEITIVKIENLLSS--GKVDLDLIDSEVTRIKELKFSIES 918

Qy   410 IRKVYTEADDLFEN--------KIGGGTVDFINKEIRDPSKALIRKVST--------- 452
         :|: :: :: |::         |  ::||||::|:||    ||::|::
Db   919 LESYYLEKIDEFRNGAIYSDELLQDIMNHFNKETRELEENLSKKFAAVLNNSEEFVKEV 978

Qy   453 DNLLE 457
```

```
RESULT 23

Db   979 DSLLQ 983
         :|:|:

G97085
sensory transduction histidine kinase (with HAMP domain) [imported] - Clostridium acetob...
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: G97085
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
  J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
R;Reference number: A96900; MUID:21359325; PMID:21359325
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: G97085
A;Accession: G97085
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-473 <KUR>
A;Cross-references: GB;AE001437; PIDN:AAK79474.1; PID:g15024454; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
A;Genetics:
A;Gene: CAC1507

Query Match       4.9%;  Score 116.5;  DB 2;  Length 473;
Best Local Similarity    21.8%;  Pred. No. 3.2;
Matches    86;  Conservative  62;  Mismatches 120;  Indels 127;  Gaps  20;

Qy   86 EAKYQPLTLPDPYQLEAAFILFKESDANPANSTEKRF--------------WMRFRGKNHSYFH 136
Db   16 EAFFVPSSI---YLINSSFKSNLQDEISSGISEEKRFSSLIQANLTFFKIKRGK------Y 67

Qy  137 DLVFNLLEKNVTRDADA--TDIENFASRYLYMAFLYKYTYTNVDEFGASFFNKLSFTTG 193
Db   68 DVSYNLTKE----ELDVFINTYLSDFKDQKYYIDV-------IDEKGNSVFNDLS---- 111

Qy  194 LFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSSYKDYMDTQIPALPKFAKRFSLMV 253
Db  112 -----NKLPKKRPELDLNKNQLKYILRDING--KTY------------------LFI 143

Qy  254 VQRLLATVAGYVDFPWYKKWYMK--------------LK-NFMVNRVFIPTKFKFFNKE 296
Db  144 VSKL-----KLDNSYYKFSYIKDVSKIYDNRKYLLNVLLKLNIFIIILIIVMLLISKF 197

Qy  297 IREPSKALKEKVSFDTKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIG 356
Db  198 IVKPINVMIKST----------QKIAEGN---FN------ERVSGVRND---EIG 230

Qy  357 QGTVDFINNEIRDPSKALIRKVSTGCAEDLFENKIGQGTVDFINNEIRDPSKALIRKVYTE 416
Db  231 QLSKNF--NYMADVIEDKIKELKTSSED------KQRFIDDLTHEIRTPLTSII---GY 278

Qy  417 ADDLFENKIGQGTVDFINKEIRDPSKALIRKVSTE 451
Db  279 ADFLRTAKYDEKTLFSSLNYIYDEGKRL-QKLSSK 312


RESULT 24

G82925
hypothetical protein UU162 [imported] - Ureaplasma urealyticum
C;Species: Ureaplasma urealyticum
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: G82925
R;Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A;Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min...
A;Reference number: A82870
A;Accession: G82925
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-457 <GLA>
A;Cross-references: GB;AE002116; GB;AF222894; NID:g6899118; PIDN:AAF30568.1; GSPDB:GN001...
```

```
C;Genetics:
A;Gene: UU162
A;Genetic code: SGC3

Query Match       4.9%;  Score 116;  DB 2;  Length 457;
Best Local Similarity    19.4%;  Pred. No. 3.3;
Matches    83;  Conservative  61;  Mismatches 133;  Indels 150;  Gaps  19;

Qy   78 LQSLKYPLEAKYQPLTLPDPYQLEAAFILFKES---DANPANSTEKRFWMRFRGKNHSY 134
Db   25 LEDNQLQLERKKQALT----------TFILRAQSMIDDCVASNG------GAGFIYSW 66

Qy  135 FHDLVFNLLEKNVTRDADAT----------------------------------- 154
Db   67 FDDLKTTDYQKTKFKEAIATLTNYLWLQGHFIDKDYTSAIGDGLSLSFKSENADFEKKRL 126

Qy  155 DIENFASRYLYMATLYKT-----------YTNVDEFGASFFNKLSFTTGLFGWGIKRALK 204
Db  127 DIIN----KLRVCGLYKKSKFANFNDDGIYVNDELKINAIN------NVHEELR 171

Qy  205 Q------IIRSNLPLDIGTEHSVSRLQHITSS-----YKDYM--DTQIP---ALPKFA 246
Db  172 DHYISKKNNPVVNNNFDLN--HHNINNVSSINASGNLATLNGFILNDCTLPNDLTTDKSN 228

Qy  247 KRFSLMVVQRLLATVAGYVDFPWYKK--WYMKLKNFMVNRVFIPTKKFFNKEIREPSKAL 304
Db  229 KVYD--------NAIANYVDIHDVKNLAWYGQNNNNAITR--------QELYEYCNKI 270

Qy  305 KEKVSFDTKDLFENKIGQGTVDFDFFNK---------------------EIRDPSKALKEKVSN 345
Db  271 NSKMYWDA--IIDYKAGELIVDLDEKTNTIKVYKALVNNGVSTHNIVDPSKQINENVWEL 328

Qy  346 -DAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGCAEDLFENKIGQGTVDFINNEIRD 404
Db  329 INIQQIDLNKIREGVQDAVDNKVNEVKKAIDFKENTFTKNL-ENKVNEAKQAVLSSYEQD 387

Qy  405 PSKALIR 411
Db  388 KASGVFK 394


RESULT 25

E81339
probable restriction /modification enzyme Cj0690c [imported] - Campylobacter jejuni (
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C;Accession: E81339
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
  C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Whitehead, S.; VanVliet, A.;
nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912
A;Accession: E81339
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1250 <PAR>
A;Cross-references: GB;AL139076; GB;AL111168; NID:g6968128; PIDN:CAB72964.1; PID:g696
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0690c

Query Match       4.9%;  Score 116;  DB 2;  Length 1250;
Best Local Similarity    19.9%;  Pred. No. 12;
Matches    81;  Conservative  56;  Mismatches 145;  Indels 126;  Gaps  15;

Qy  134 YFHDLVFNLLEKNVTRDADADATDIENFASRYLYMATLYKYTYTNVDEFGASFF-------- 185
Db  430 YFDNYDFKILEK------SKKVEKYT--FYRKGQIYLKNTSNSRKASASFYTPQSIANF 480

Qy  186 -------------NKLSFT-----TGLFGWGIKRALKQIIRSNLP------LDIGTEHS 220
Db  481 LIQSALKDKLINNENILKFKILDNACGSGHFLVGVLNAITHIVLSDFDHFTNLKELYEEK 540
```

```
Qy   221  VSRLQHTTSSYKDYMDTQIPALPKFAKRFSLMVVQRLLATVAGYVDFPWYKKWYMKLKNF  280
               :|| ||  |
Db   541  ENILNHIKDFVQDYEVDSDIIKRL---------LLKRIIYGVDLNPFSIELTKLSLW  589

Qy   281  MVNRVFIPTKKFFNKEIREPSKALKEKVSTDDTDKDLFENKIGQGTVDFINKEIRDPSKALK  340
          |  : |
Db   590  IDSFITTPLSFIEHHIKCGNALINSNLS-DEKDL----IKQNSSNLFTNSITQEFFILQ  644

Qy   341  ----EKVSN---------DAKDLFENKIGQ-------GTVDFINNEIRDPSKALI  375
             |:||
Db   645  EVFEKLDNLKDTNEEQIKQSKQIYQNEITPKLNKLNLYLNYINTLHFVNKELQILKAL-  703

Qy   376  RKVSTGAEDLFENKIGQGTVDFINNEIRDPSKALIRKVYTEAD------DLFENKIG  426
               : |            | |
Db   704  -----SQD----------DIQNLSQNEQAKAIISKYQKEFNFNYELEFPEIVENQVF  746

Qy   427  QG------------TVDFINKEIRDPSKALIRKVSTEADNLLEK  458
                        |||| |
Db   747  KGFDIIIGNPPWDKTKFSDSDFFPQYKSDYRSLIASKKKEIQDNLLAK  794
```

RESULT 26
H71931
hypothetical protein jhp0440 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 28-Jul-2000
C;Accession: H71931
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
;Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Reference number: A71800; MUID:99120557
A;Molecule type: DNA
A;Residues: 1-912 <ARN>
A;Status: preliminary
A;Cross-references: GB:AE001478; GB:AE001439; NID:g4154971; PIDN:AAD06018.1; PID:g415497
A;Experimental source: strain J99
A;Gene: jhp0440
C;Genetics:
A;Gene: jhp0440
C;Superfamily: Helicobacter pylori hypothetical protein jhp0440

Query Match   4.9%; Score 115; DB 2; Length 912;
Best Local Similarity 20.5%; Pred. No. 9.4;
Matches 105; Conservative 51; Mismatches 167; Indels 188; Gaps 24;

```
Qy    98  YQLEAAFILFKESDANPA---NSTEKRFWMRFRGKHHSFYFHDLVFNLLEKNVTRDADAT  154
          ||   : :
Db   312  YDLDYAIV--KAKDLKPSFTTCGTQKRTDM-----NEEQIKSIAENFDPKIFGSGGFE  363

Qy   155  DIE---------------NFA---SRYLYMATL--YYKYTYNVDE------FGA  182
          |||             |||      :|
Db   364  DLPIILHDGQVIAGNHRIAGMLNFTPKSRYIYNKAIKEYYRIDLKPDELLVRVPHQRLDN  423

Qy   183  SFFNKL---SFTTGLFGWGIKRA-------------LKQIIRSNLPLDI  215
                   :||
Db   424  TEINNLAASSNQGRFNSESDHAIAIVLSHYEARKLKELDKKLDADSIYSLKNIVAKNLNFDK  483

Qy   216  GTEHSVSRLQHTTSSYKDYMDTQIPALPKFAKRFSLMVVQRLLATVAGYVDFPWYKKWYM  275
          |  |||
Db   484  ATHPNVG-------------DSNLALLMYNMPRTKTQGIELL---  521

Qy   276  KLKNF-MVNRVFIPTKKFFNKEIRE-----PSKALK-------  314
               :|:
Db   522  DIKSYEKVKKMFVDNAGSFHNLIHDMNFPNVSLNAYLSDIMDRSFANLKNYPSTSESLKD  581

Qy   315  LFENKIGQGTVDFFENK------------EIRDPSKALKEKVSND--AKDLFE  352
          |:| |  |
Db   582  LSEKFYKTSSLDMFEKSDQSASDISEILGGAIARFARFDDPSKALFESLSKSDNIKGLKD  641

Qy   353  NKIGQGTVDFIN---NEIRD-------------PSKA  373
            |   :   |  |  :|
Db   642  YKIADITKDMFNPDSKEFKDIDIYDFTHYLLMVNREPNENNPVLNRLIQAVKGMQKESEK  701

Qy   374  LIRKVSTGAEDLFENKIGQGTVD-FINNEIRDPSKALIR-KVYTEADD-LF-ENKIG  426
          |:| ||
Db   702  GIKEVSKQAAEETEKGIKRDFSDTNLSNNEI---KALLNNAKIPISGRDAVIFGKNNLN  757

Qy   427  QGTVDFINKEIRDPSKALIRKVSTEADNLLE  457
          |:||
Db   758  PEIVEFIHK---NHKRMAIEKASNKEIELLQ  785
```

RESULT 27
S49394
HsdRI protein - Mycoplasma pulmonis
C;Species: Mycoplasma pulmonis
A;Molecule type: DNA
A;Residues: 1,986 <DYB>
C;Date: 29-Nov-1995 #sequence_revision 01-Mar-1996 #text_change 07-Dec-1999
A;Cross-references: GB:L25415; NID:g496153; PIDN:AAA65632.1; PID:g496157
C;Accession: S49394
A;Accession: S49394
A;Status: preliminary
A;Title: Regulation of a restriction and modification system via DNA inversion in Myc
R;Dybvig, K.; Yu, H.
Mol. Microbiol. 12, 547-560, 1994
A;Reference number: S49391; MUID:95020589
A;Genetic code: SGC3

Query Match   4.9%; Score 115; DB 2; Length 986;
Best Local Similarity 23.1%; Pred. No. 10;
Matches 81; Conservative 67; Mismatches 142; Indels 60; Gaps 17;

```
Qy   140  FNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEFGASFEN--KLSFTTGLFGW  197
          |:|
Db   648  FSLEEQSIN--DAFKIYANSSDKELQQLVYGEKYEQVVEDFINFWNSLKISFS------  698

Qy   198  GIKRALKQIIRSNLPLDIGTEH--SVSRLQHTTSSYK------------DYMDTQIPALP  243
          |                      :::
Db   699  NIYDEKNNEIFRNISLENKKKYLKNLSQVSNIFSSLKTFKEYGKNEKISDFSLEQLNQQ  758

Qy   244  KFA-----KRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNPMVNRVFIPTKFFNKEI-R  298
          |:|
Db   759  KWANEIKKNLSTNEKEKISYEVLNSIDISNIKFAV--KEMIIDEIYLENLLFPNKKISK  815

Qy   299  EPSKAL--KEKVSTDTK--DLFENKIGQGTVDFINNEIRDPSKALKEKVSNDKADLFENK  354
          :||
Db   816  YPNNRLTYEDTLSEIDKHIQLIKNNYNQGKI---NQKEYIFLLLVQKWKNEIKNEFIKK  872

Qy   355  IGQGTVDFINNEIRDPSKALIR---KVSTGAEDLFENKIGQGTVDFINNEIRDPSKALI  410
          |:|
Db   873  -DKSLD--EKEFIDYGKRILKSVFQKVKNQIEANWLEKILKEYHGINDQIRKDWKKRI  928

Qy   411  RKVYTEADDLFENKIGQGTVDFINK---EIRDPSKALIRKVSTEADNLLE  457
          |:|
Db   929  ND--KDLDDIEKS------EFIKKWSRRSKEVDKDIIDKLSIEYKESIE  969
```

RESULT 28
S38173
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKR095w; protein YKR415
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C;Accession: S38173; S40647; S31207
R;Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantale
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38158
A;Accession: S38173
A;Molecule type: DNA
A;Residues: 1-1875 <BAL>
A;Cross-references: EMBL:Z28320; NID:g486586; PID:g486587; MIPS:YKR095w
A;Experimental source: strain S288C

R:Bou, G.; Esteban, P.F.; Baladrón, V.; González, G.A.; Cantalejo, J.G.; Remacha, M.; Ji...
Yeast 9, 1349-1354, 1993
A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromo...
A:Reference number: S40647
A:Accession: S40644; MUID:94205265
A:Molecule type: DNA
A:Residues: 1-1875 <BOU>
A:Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A:Experimental source: strain S288C
A:Molecule type: DNA
R:Kölling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A:Title: A new yeast gene with a myosin-like heptad repeat structure.
A:Reference number: S31207
A:Accession: S31207; MUID:93247549
A:Molecule type: DNA
A:Residues: 1-300,'A',302-1875 <KOE>
A:Cross-references: EMBL:L01992; NID:g171958; PIDN:AAA34783.1; PID:g171959
A:Genetics:
A:Cross-references: SGD:S0001803; MIPS:YKR095w
A:Gene: SGD:MLP1
A:Map position: 11R

```
Query Match    4.9%;  Score 115;  DB 2;  Length 1875;
Best Local Similarity   20.0%;  Pred. No. 24;
Matches  109;  Conservative   92;  Mismatches  211;  Indels  134;  Gaps  24;

Qy       4  SDSVGDVTKLLAASESVDSAANAYMNINSDMSDYLSAVSDNFAERICSQVPKGSNCSASV   63
Db     523  SDSQKVTERLV-------EFKNIIQLQEKNAELIKVVR-NLADKLESKEKKSKQSLQKI   574

Qy      64  SAYMSRCAKQDCLTLQSLKYPLEAXYQPLTLPDPYQLEAAFILFKESDANPANSTEKRFW   123
Db     575  ESETVNEAKEAIITLKSEKMDLESRIEEL--------QKELEELKTSVPNEDASYSNVTIKQLT   630

Qy     124  MRFR---------------------RGKNHSYFHDLVFNLLE-------   144
Db     631  ETKRDLESQVQDLQTRISQITRESTENMSLLNKEIQDLYDSKSDISIKLGKEKSSRILAE   690

Qy     145  ---KNVTRDADATDIEN--FASRLYMATLIY-------KTYTNVDEFGASFFNKLSFT-TGL   194
Db     691  ERFKLLSNTLDLTKAENDQLRKFDYLQNTILKQDSKTHETLNEY-VSCKSKLSIVETEL   749

Qy     195  FGWGIKRALKQIIRSNLPLDI---------------GTEHSVSRLQHITSSYKDYM-----   235
Db     750  LNLKEEQKLRVHLERNLKQELNKLSPEKDLSRIMVTQLQTLQKEREDLLEETRKSCQKKI   809

Qy     236  DTQIPALPKFAKRFSLM--VVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFF   293
Db     810  DELEDALSELKKETSQKDHHIKQLEEDNNSNIE--WYQNKIEALKDYESVITSVDSK--   865

Qy     294  NKEIREPSKALKEKVYSTDTKDLFENKIGQGTVDFENKEIRDPS---KAL-KEKVS-----   344
Db     866  ----QTIDIEKLQYKVKSLEKEIEEDKIRLHTYNVMDETINDSLRELEKSKINLIDAYS   921

Qy     345  -NDAKDLFE--------NKIGQGTVDFIN--NEIRDPSKALIRKVSTGAEDLFENKI   390
Db     922  QIKEYKDLYETTSQSLQQTNSKLDESFKDFTNQIKNLTDEKTSLEDKISLLKEQMEN--   978

Qy     391  GGGTVDFINNEIRDPSKALIRKVYFEADDLFENKIGQGTVDFINKE---IRDPSKALIRK   447
Db     979  -------LNNELDLQKKGMEKE---KAD--FKKRI--SILQNNKEVEAVKSEYESKLSK  1024

Qy     448  VSTEAD   453
Db    1025  IQNDLD  1030
```

RESULT 29
E89897
conserved hypothetical protein SA1085 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: E89897
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O...
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Savano, T.; Inoue, R.; Kaito, C.; Sekimizu, K...
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-868 <KUR>
A:Cross-references: GB:BA000018; PIDN:g137010042; PIDN:BAB42337.1; GSPDB:GN00149
A:Experimental source: strain N315
A:Gene: SA1085

```
Query Match    4.98%;  Score 114.5;  DB 2;  Length 868;
Best Local Similarity   19.9%;  Pred. No. 9.5;
Matches   62;  Conservative   48;  Mismatches  125;  Indels   77;  Gaps  12;

Qy     113  NPANSTEKRFWMRRPRGKNHSYFHDLVFNLLEK---NVTRDADATDIENFASRLYMATL   169
Db     451  NNKNMAIKPYQQSLSTLKQHDYHSNYVNQLIKKINQNATGPFNRID--YMSDYALNSPF   507

Qy     170  YKYTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLD--------IGTEHSVS   222
Db     508  IY----------HYNGISLYSSIFNGDILKYDKTLQINMPIDKNSTYRLLGNRQNLL   555

Qy     223  RLQHITSSYKDYMDTQIPALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMV   282
Db     556  SLWNVNDRIRVNHDDNLP-----YRKIN-----SEHKDNK--VRWIHSKNTIHY   598

Qy     283  NRVFIPTKKFFNKEIREP----SKALIKEKVSTDTKDL-----FENKIGQGTVDFFNKEI   332
Db     599  PSAHITNKVFSNKELKSPLDKEQAMLQGIVSNNTKDVNTHFKANKNLLSDSTIKLNSSAW   658

Qy     333  RDPSK-------------ALKEKVSNDAKDLFENKIGQGTVDFFNEIRDPSKALIRKV   378
Db     659  QSPTKHLLQVKQNNGGLTVQLPKSVSNQFKDLYFE-----MDLELLSPDKAHDVKV   709

Qy     379  STGAEDLFENKI   390
Db     710  NEYTQE--RNKL  719
```

RESULT 30
F90561
hypothetical protein MPU_3980 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: F90561
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, ...
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p...
A:Reference number: A95512; MUID:21267165; PMID:11353084
A:Accession: F90561
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-652 <KUR>
A:Cross-references: GB:AL445566; PIDN:g14089812; PIDN:CAC13571.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
A:Genetics:
A:Gene: MPU_3980
A:Genetic code: SGC3

```
Query Match    4.8%;  Score 114;  DB 2;  Length 652;
Best Local Similarity   19.6%;  Pred. No. 7.1;
Matches   96;  Conservative   75;  Mismatches  180;  Indels  140;  Gaps  21;

Qy      76  LTLQSLKYPLEAXYQPLTLPDPYQLEAAFILFKESDANPANSTEK--RFWMRRPRGKNHSY   134
Db     148  IALSSLLERYEEKVKLIYIDPYYFEDK---KEENTFLYNSNFKLSTWLSFMK-----   197

Qy     135  FHDLVFNLLE--KNVTRDADATDIENFASRLYMATLYKTYTNVDEFGASFFNKLSFTT   192
```

RESULT 31

```
A26655
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
C;Contains: myosin ATPase (EC 3.6.1.32)
C;Species: Dictyostelium discoideum
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 02-Feb-2001
C;Accession: A26655; A24728; S00250
R;Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A;Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium disco
A;Reference number: A26655; MUID:87092266
A;Accession: A26655
A;Molecule type: DNA
A;Residues: 1-2116 <WAR>
R;Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A;Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium
A;Reference number: S00250; MUID:88112226
A;Accession: S00250
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
R;DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A;Reference number: A24728; MUID:86016788
A;Accession: A24728
A;Molecule type: mRNA
A;Residues: 2035-2116 <DEL>
A;Cross-references: GB:M14628; GB:M11938; NID:g167834; PIDN:AAA33227.1; PID:g167835
A;Molecule type: DNA
A;Accession: A26655
A;Reference number: A26655; MUID:87092266
A;Molecule type: mRNA
A;Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phos
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
F;1-818/Domain: myosin motor domain homology <HED>
F;89-747/Domain: globular head <HED>
F;179-186/Region: nucleotide-binding motif A (P-loop)
F;819-2116/Domain: alpha-helical rod <ROD>

Query Match             4.8%;  Score 113.5;  DB 2;  Length 2116;
Best Local Similarity  20.0%;  Pred. No. 35;
        Matches   91;  Conservative   77;  Mismatches  166;  Indels  121;  Gaps  23;
```

```
Qy    198 -----------NRLEITKMLLREDGFIFVQISDDGFAVLKVIMDEIFGN-KKIINTIVVKTKASS 250
                      :||| ||: || |  :    |  : :||       |      ||:  ::
Db    606 EEKKKALIKLV 616
```

```
Qy    193 GLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALPKFA------ 246
              ::|||:: |  ||::||: :  : ||   :   :|   :|||   |   :|
Db    251 GASGGGEDKKLKKNI--EVILVVGKENALKAQHIEIPLNKYINNKRNEKKNFAYTNVML 308
```

```
Qy    247 -----------KRFSLMVVQRLLATVAGYVDTPWYKKWYMKL----- 277
                      ::   :  ||   || || |   :  ::
Db    309 DIGEEFYIGETKDGYGETIKLYQMKNFKTKSIKE-LAKEENCSDEEIYNKYIDKIYTTEN 367
```

```
Qy    278 -KNFMVNRV-------------------FIP--------TKKFFNKEIREPSKALREKVSTDK 313
            :|| :||::                   |||         |||||  |   | |:| |
Db    368 AQTSIRNVRENVSDDVELVIARYVPISGKNKGKLTDIGFIGKTRRLISYLKETIIFKNN 427
```

```
Qy    314 DLF-ENKIG---------OGTVDFFNKEIRDPSKALKE--KVSNDAKDL-FENKIG 356
            :|: |||:           :    |||||| ||:|||:    :: |||:  |||:|
Db    428 SIFKQEKVGTLWSDLSWSSVKFEGKIDFGSGQ--KPEKLIERIKLGTEKNDLVDFHLG 485
```

```
Qy    357 OGTVDFINNE---------- 402
           ::     |:|
Db    486 SGTTTAVAHKMGRKYIGIGEOMDYIODIITIERMKKVIDGEQGISKNVDWKGGGSFVYFEL 545
```

```
Qy    403 RDPSKALIRKVYT------------------I 437
          |||:|||:||:                    :
Db    546 LENANELIEKVYSSTYIPERERAETISTIKNEIYNDERIVPYITKEELKNADKEFENLNT 605
```

```
Qy    438 RDPSKALIRKV 448
          |||:||||| |
Db    606 EEKKKALIKLV 616
```

RESULT 32

```
A56577
microtubule-associated protein MAP 1B - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Feb-1997
C;Accession: A56577
R;Zauner, W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.
Eur. J. Cell Biol. 57, 66-74, 1992
A;Title: Identification of two distinct microtubule binding domains on recombinant ra
A;Reference number: A56577; MUID:92347374
A;Accession: A56577
A;Molecule type: preliminary
A;Residues: 1-2364 <ZAU>
A;Cross-references: GB:X60550
A;Experimental source: brain
A;Note: nucleotide sequence not given; conceptual translation not complete
C;Superfamily: microtubule-associated protein MAP1B

Query Match             4.8%;  Score 113.5;  DB 2;  Length 2364;
Best Local Similarity  20.7%;  Pred. No. 40;
        Matches  100;  Conservative   76;  Mismatches  176;  Indels  131;  Gaps  21;
```

```
Qy     42 SDNFAER-------------ICSQVPKGSNCSASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLP 95
          :||:             ::   |||   : :|  :| | ||  | |   ::    | ||  ||:|
Db    604 SDNVVTKLFNDPNIASRAKKGANFIIVAAQY----KEQLASLMATLETTNPHFVRCIIP 658
```

```
Qy     96 D-----PYQLEAAFILFKESDANPANSTEKRFWMRFRRGKNHSYFHDLV--FNLLEKNVTR 149
            :     |  |  |::  ||:| : :|:||   | |||| :::: |    |  |:||:|
Db    659 NNKQLPAKLEDKVVL----DQLRCNGVLEGIRITRKGFPNRIIYADFVKRYYLLAPNVPR 714
```

```
Qy    150 DAD----ATDIENFASRYLYMATLYYKTYTNVD----EFGASFFNKLSFTTGLFGWGIKR 201
          :|:          |  : |: :|    | ::           | :|||:::
Db    715 DAEDSQKATD------AVL--KHLNIDPEQYRFG---ITKIFFRAGQLA-RIEE 756
```

```
Qy    202 ALKQIIRSNL-PLDIGTEHSVSRLQHITSSYKDYMDTQIPALPKFAKRFSLMVVQRLLAT 260
          |::|:|||   :|::|    : :  :|| : |||  |   |:   |:     :: :|
Db    757 AREQRISEIIKAIQAATRGWIAR-------KVYKQARHTVAA--------RIIQONLRA 801
```

```
Qy    261 VAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREPSKAL-------------KEK 307
          |||||||||:::  :||||| | || |||||| |||||||||||              |:|
Db    802 YIDFKSWPWW-KLFSKARPLL----KRNFEKIEKEREILELKSNLTDSTTQKDK 853
```

```
Qy    308 VSTDTKDLFENKIGQGTVDFDFNKEIRDPSKALKE--KVSNDAKDLFENKIGQGTVDFINN 365
          :|:  :|||||| ||:| |||||||| |||:|||:    :: ||| |||:| ||:||::|
Db    854 LEKSLKDTESNVL-------DLQRQLKAEKEETLKAMYDSKDALE--LQNQKRSVEEKVRDLEEELQEEEOK 940
```

```
Qy    366 EIRDPSKALIRKVSTGAEDLFENKIGQGTVDFINNEIRDPSKALIRKVYTEADDLFENKI 425
          ||||||||:|||||: |||| |||:| ||:||::||||||||||:||||: |  | |||
Db    891 ------AQKRELEIRVEDM-ESELDEKKLALEN--LQNQKRSVEEKVRDLEEELQEEEOK 940
```

```
Qy    426 GOGTVDFINKEIRDPSKALIRKVSTEADNL--LEK 458
          | || ||:||:|||||||:|||||:    |   :||
Db    941 LRNTLEKLIKKKYEEELLEEMKRVNDGQSDTISRLEK 975
```

```
Qy     19 ESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSASVSAYMSRCAKQDCLTL 78
          ||:|:  | |: :|   :  : :   | |:|||   : :|| :| | ||| | |   :|
Db    290 QYLN-KLSMKPEPLFRSVGNAIEPV-ILFQKMGVGKLKMYVLNPVKSSKEMOYFMQQWTG 347
```

```
Qy     79 QSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDA------------NPANST-EKRFWMRFRRG 129
          ::|:|  :||:|||||   ||| | |::||||           |||| :||:|
Db    239 ESQGSTSNSDWMKNLISPDLGVVFLNVPENLKNPEP-----NIKMKRSTEEACFTL 289
```

```
Qy    130 KNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEFGASFFNKLS 189
          ::: |||:|  | :|::|   |:  |:|| :|||:| |:|   | :::
Db    348 TNKDKRAELILPNGQEVDIP----------ISYLA 371
```

```
Qy    190 FTTGLFGWGIKRALKQIIRSNLPLDIGTEHSV----SRLQHITSSYKDYMDTQIPALPKF 245
          :: |||||||||||:||||| :|::|   |||     |||||:|||||| |   | |
Db    372 SVSSLIVWHPANPAEKIRVLFPGN-STQYNILEGLEKLKHL----DFL------- 415
```

```
Qy    246 AKRFSLMVVQRLLAT--VAGYVDTPWYKKWYMKLKNFMVNRVFI--PTKKFFNKEIREP 300
```

```
Db  416  ------------KQPLATGKDLTGQVSTPPVKQ--VKLKQRADSRESLKPATKPLSSSKSVRKE 464
         :::  |||          ::|:  ||             :|:   |: ::     :|:
Qy  301  SK-----------------ALKEK--VSTDTKDLFENKIGQGTVDFFNKEIRDPSKA-L 339
            |               |||    ||   :|  | ||||||||||||||||| |
Db  465  SKEEAPEATKASQVEKTPKVESKEKVIVKKDKPGKVESKPSVTEKEVPSKEEQSPVKAEV 524
Qy  340  KEKVSNDAKDLFENKIGQGTVDFINNEIR-------DPSKALIRKYSTGAEDLFENKI 390
         |||     |:|||||||||||||||||||       |||||||||||| | ||  |||
Db  525  AEKAATESKP-------KVTKDKV--VKKEIKTKPEEKKEEKPKKEVAKKEDKTPLKKDKPK 578
Qy  391  GQGTVDFINNEIRDPSKALIRKVYTEADDLFENK--IGQGTVDFFNKEIRDPSKALIRKV 448
         ||||||||||| :|||     | | |  | |||    :|||||||||||||||||||||
Db  579  KEEAKKEIKKEIKKEEKKELKKEVKKETPLKDAKKEVKDEKEVKKEEKPKKE-IKKI 637
Qy  449  STE 451
         |:|
Db  638  SKD 640
```

RESULT 33
A85176
hypothetical protein AT4g15890 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: A85176
A;Accession: A85176
A;Molecule type: DNA
A;Residues: 1-1314 <STO>
A;Cross-references: GB;NC_001268; NID:g7268337; PIDN:CAB78631.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g15890
A;Map position: 4
A;Status: preliminary
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488

Query Match   4.8%;   Score 112.5;   DB 2;   Length 1314;
Best Local Similarity 20.2%;
Matches 99;   Conservative 89;   Mismatches 190;   Indels 113;   Gaps 27;

```
Qy   14  LLAASESVDSAANAYMINSDMSDYLSAVSDNFAERI---------------------SQVP 54
         |:| | :||  | | ||:||||||||||||||||                       | |
Db   10  ILAGIEGGDDESDYHELVTDLKSLLDTDDDEINRFVGSLSSMASSFLRCISAAAMDSPVE 69
Qy   55  KGSNCSASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANP 114
         :|:|                  :|  | :| :|      : :|  |  ||:|: |||
Db   70  SGRLAILASDAYLLSLLSTNC------PVTFFSPIAFLSLLGSIRRYLKRRDSAGQ 121
Qy  115  ANSTEKRFWMRPRGRKNHSYFHDLVFNLLEKNV-TRDADATDIENFASRYLYMATLYYKT 173
          |||:|: |  :: |:     |:  : | ::|  |||  |  : |  :|: :: | |:|
Db  122  GSNSQREKGNNKKRGRG-----KRNLGYEDGEETEEGGFDAKLMFIV----- 163
Qy  174  YTNVDEFGA--SFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHS----VSRLQHI 227
         |:| :|     | |||||  |||||| || : |  |: |    ||         |: |
Db  164  ---LEKLGSVLSFVHLDRFPDSL--KSLVQTV-SEIPL-LALEHSGVLNYDRLMEM 212
Qy  228  TSSY------KDYMDTQIPAL---PKPAKRFSLMVVQRLLATVAGYVDTP 268
          |::       |   |  |      |:| : ||:::|::| : | |  |
Db  213  CGKILGGVVLNSDHGDMALTAAEISKSLTPLLMGKHQARSFALGFVSRKLMSLA--KDNP 270
Qy  269  WYKKWYMKLKNFMVNRVFIPTKK-----FFNKEIREPSKALKEKVSTDTKDLFENKIGQGT 324
         | |:|  | |  : :: :| |       |:|| |  |||:|  |:| |||||||||||||
Db  271  ELKKVVSNLPKFLVHKA--PEKAEPRGFAVEAVLEIVKAMEVEGQSEFVD-FVMKMCGGK 327
Qy  325  VDFFNKEIRDPSKALKEKVSNDAKDL----FENKIGQGTVDFINNEIRDPSKALIRKVS 379
         |||:||||||||||||:|:||  |       |||||||||:|||||||||||||||||:
Db  328  SNFRVLAV-DIPLLISSLGNPLGDISSENGLKDSWGLGCIDALVQRCSDTS-ALIR- 382
Qy  380  TGAEDLFENKIGQGTVDFINNEIRDPSKALIRKYVYTEADDLFENKIGQGTV-DFINKEIR 438
         |||||||||||||||||||||||||||||||||:||:||||||||||||||| :|||:||
```

```
Db  383  --ARAL--SNLAQ-VVEFLSGDER--SRSILKQALGFNGETSE--GKGAVTDLLKKRCV 432
          | ||     | |   | :: ||    : ||: |  :| :|     |  | |||||:|
Qy  439  DPSKALIRKVS 449
         ||||||||||:
Db  433  D-EKAAVRRAA 442
```

RESULT 34
T38077
hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38077
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z21767
A;Accession: T38077
A;Molecule type: DNA
A;Residues: 1-1957 <CON>
A;Cross-references: EMBL:Z70690; PIDN:CAA94624.1; GSPDB:GN00066; SPDB:SPACIF3.06c
C;Genetics:
A;Genetics:
A;Experimental source: strain 972h-; cosmid c1F3
A;Gene: SPDB:SPAC1F3.06c
A;Map position: 1
A;Residues: 1-1957 <CON>
A;Status: preliminary; translated from GB/EMBL/DDBJ

Query Match   4.7%;   Score 112;   DB 2;   Length 1957;
Best Local Similarity 19.2%;
Matches 107;   Conservative 94;   Mismatches 204;   Indels 152;   Gaps 27;

```
Qy    9  DVTKTLLAASESVDSAANAY-MINSDMSDYLSAVSDNFA------------ERICSQVPKGSN 58
         |: |  :||:|||||| |   :||:|||||||||||||            ||||:| |||:|
Db    5  DLSRQLLTVTEKLDDKKEKDYEKIKEDVSSIKASLAEEQASNKSLRGEQRL-EKLLVSSN 263
Qy   59  CSASVSAYMSRCAKQDCLTLQ-------------SLKYPLEAKYQPLTLPDPYQLE 101
         :|                                |||| ||||||||||||||||
Db  264  KTVSTLRQTENSLRAECKTLQEKLEKCAINEEDSKLLEELKHNV-ANYSDAIVHKDKLIE 322
Qy  102  AAFILFKESDANPANSTEKRFWMRFRRGK------------------NHSYFHDLVFNLLE 144
          |:||||||||| ||||||| |  :: ||                  |:|||||:||||
Db  323  DLSTRISEFD----NLKSERDTLSIKNEKLEKLLRNTIGSLKDSRTSNSQLEEMVELKE 378
Qy  145  KNVTRDADATDIENFASRYLIYMATLYKYTVNVDEFGASFFNKLSFTTGLFGWGIKRALK 204
         |||:| | ||:||  :|: | | |:||:|| | |||:|||||||||||||:|||||||||
Db  379  SNRIHSQLTDAESKLSSF---EQENKSLKGSIDEYQ---NNLSSKDRM----VKQVSS 427
Qy  205  QI--IRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALPKFAK--RFSLMVVQRLLAT 260
         ||  ::|  |     ||   | |: : | ||:|| ||  | |     ||:|::|::|:|
Db  428  OLEEARSSL------AHATGKLAEINSE-RDFQNKKIKDFEKIEQDLRACLNSSSNELKE 480
Qy  261  VAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREPSKALKEKVSTDTK--DLFEN 318
         | |:  |  | ||  | |  :  ::  | ||  |:: :  |||||||:||:|    |||:
Db  481  KSALID-----KKDQELNNL---REQIKEQKKVSESTQSSLQSLQRDILNEKKHHEVVES 532
Qy  319  KI-------GQGTVDFFNKE------IRDPSKALKEKVSN 345
         |:       |||||||:||       |||||||:|:|||
Db  533  OLNELKGELQTEISNSEHLSSQLSTLAAEKEAAVATNNELSESKNSLQTLCNAFQEKLAK 592
Qy  346  DAKDLFEN-----------KIGQGTVDFFNKE-----EIRDPSKALIRKVSTGAEDLFE 387
         | |||||            ||||||||||||     |||||||||||||||||||||:|
Db  593  SVMQLKENEQNFSSLDTSPKKLNESHQELENNHQTITKQLKDTSSKLLQQLOLERAN--FE 650
Qy  388  NKIGQGTVDFINNEIR--------DPSKALIRKVYTEADDLFENKIGQGTVDFINKEIRDP 440
         |||||||||||:|||         |||||||||:||:||||||||||||||||||:|||||
Db  651  QK--ESTLSDENDLRTKLLKLEESNKSLIKK-QEDVDSLEKN-----IQTLKEDLRKS 701
Qy  441  SKALIRKVSTEADNLLE 457
         |||||||||:||| |||
Db  702  EEAL-RFSKLEAKNLRE 717
```

RESULT 35
T50382

RESULT 36

G90576

C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C;Accession: T50382
R;Saunders, D.; Harris, D.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
R;Reference number: Z25035
A;Accession: T50382
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-566 <SAU>
A;Cross-references: EMBL:AL133306; PIDN:CAB62095.1; GSPDB:GN00067; SPDB:SPBC902.01c
A;Experimental source: strain 972h(-); cosmid c902
A;Genetics:
A;Gene: SPBC428.20c ; SPDB:SPBC902.01c
A;Map position: 2

Query Match   4.7%; Score 111.5; DB 2; Length 566;
Best Local Similarity   20.0%; Pred. No. 8.6;
Matches 101; Conservative 80; Mismatches 194; Indels 129; Gaps 23;

Qy   4 SDSVGDVTKTLLAASESVDSAANAYMINSDMDSDYLSAVSDNFAERICSQVPKG------ 56
            || | :| | || :|:: | | ::| ::: || | | ::
Db   107 SQSSRDVSSHLL--DESISNPINIPSTEVE------SSNFGQTRYDQVPENPQITDMD 156

Qy   57 ----SNCSASVSAYMSRCAKQ--------------- 100
              :: :: | |:: ||
Db   157 EGLENESSISIAHDSSRLNRSTETSSSVQHTLITEADLLSSISYVLQG------I 204

Qy   101 EAAFILFKESDANPANSTEKRFWMRFRR-GKNHSYFHDL-VFNLLEKNVTRDADATDIE- 157
            | | || | || : | : | ||::|| | :|::|: : ||
Db   157 DCLTLQSLKYPLEAKYQPLTLPDPYQL 100

Qy   158 ---NFASRYLY-MATLYTKTYTNVDEFGASFENKLSFTTGL------FGWGIKRAL 203
               |||| | | || ||| |: :||| |: | : || |:| |::|
Db   205 STEYVQFKNELALLSKRIPVQYLLQMRAISETGLLYQELKVFSNYDPSVSQSIDGDNVSK 264

Qy   204 KQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALPKFAKRFSLMVVQRLLATVAG 263
            | ::|:| | | |::| |:| : ::: :| | : | :| | |:|| ::
Db   265 AFINDQSLALQSLKSVISKELTNFLALIASLDSQIRADASLEKPMVTIRRCIANTQVAKL 324

Qy   264 YVDTPWY---KKWYMKLKNFMVNRVFIPTKKFFNKEIREPSKALKEKVSTDTKDLFENKI 320
             :| | | ::| |:::| : ||| ::|| : |: | | :::
Db   325 KLRILSSVVNDNMNQENKRLIQVVSKYXNVHGDPLIQELS------DKILTEITG 373

Qy   321 GGGTV-----DFNKEIRDP--SKALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKA 373
            ||| |: ||:|| :|| |:| :|:||:||| ||| ||||| |:|| :|
Db   374 ----PLYEMIENNIYK------GELVDPYQEFF------VKEKNGSESHD----HQ 409

Qy   374 LIRKVSTGAEDLFENKIGQGTVDFINNEIRDPSKALIRKVYTEADDLFENK----IGQGT 429
            |:||| | | :| |: ||| ||| |:||:||| :|:||:| | | | |:|
Db   410 GQGDVVWGKYFLDKELIPSFLSEELVDKIFLIGKSLNFARYGCGDFDW----AQEHYQK 465

Qy   430 VDFIN------KEIRDPSKALIR 446
            |||| | |:|| ||| |
Db   466 LVKKLSYRDPHSLETVVDKAYTESINHLV----YLMEEVFHLTDHLKAIKKYLLLGQG- 519

Qy   520 -DFVDLLMESLGNSLDQPANTLFR 542
             | ||||:|| |: |:| | | |
Db   520 -DFVDLLMESLGNSLDQPANTLFR 542

---

probable spindle pole body component, probable gamma-tubulin interacting protein, yeast

RESULT 37

A46417

A46417

N;Alternate names: nuclear import protein; protein YM924.01c; protein YM9952.11c; pr
C;Species: Saccharomyces cerevisiae
C;Date: 21-Sep-1993 #sequence_revision 03-May-1994 #text_change 06-Feb-1998
A;Accession: A46417; S53979; S59302
R;Gu, Z.; Moerschell, R.P.; Sherman, F.; Goldfarb, D.S.
Proc. Natl. Acad. Sci. U.S.A. 89, 10355-10359, 1992
A;Title: NIP1, a gene required for nuclear transport in yeast.
A;Reference number: A46417; MUID:93066237
A;Accession: A46417
A;Molecule type: DNA
A;Residues: 1-812 <GUI>
A;Cross-references: EMBL:L02899
R;Connor, R.; Churcher, C.M.
submitted to the EMBL Data Library, April 1995
A;Reference number: S53979
A;Accession: S53979
A;Molecule type: DNA
A;Residues: 571-582, 'Q', 584-640, 'K', 642, 'K', 644-812 <CON>
A;Cross-references: EMBL:Z49212; NID:g798940; PID:g798951; MIPS:YMR309c
R;Churcher, C.M.
submitted to the EMBL Data Library, September 1995
A;Reference number: S59302
A;Accession: S59302
A;Molecule type: DNA
A;Residues: 1-110, 'V', 112-582, 'Q', 584-602 <CHU>
A;Cross-references: EMBL:Z54141; NID:g1072408; PID:g984682; MIPS:YMR309c

NIP1 protein - yeast (Saccharomyces cerevisiae)

Query Match   4.7%; Score 111; DB 2; Length 842;
Best Local Similarity   20.6%; Pred. No. 15;
Matches 97; Conservative 73; Mismatches 150; Indels 152; Gaps 25;

Qy   48 RICSQVPKGSNCSASVSAYMSRCAKQDCLTLQSLKYPLEAKYQP--LTLPDPYQLEAAF- 104
            ||| ||||: :::|| | | |: |:||| ::| ||: ||| | || || |
Db   70 KIVAQNPSKGNIKVTLLPFKRVKVNKITISETNLSHKVMVDYSPFELVAPDPTKGKILYN 129

Qy   105 ----ILFKESDANPANSTEKRFWMRFRRGKNHSYFHDLVENLLEKNVTR--DADATDIE 157
                || | |: || |:| | | | :|: | :| :: |::: || |||
Db   130 KMVDIIFKNS--MPLDSESKR------NISSL-SIKYNSIAKELDKLKIESDQEFLE 177

Qy   158 NFASRYLYMATLYKTYTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGT 217
            ||||| | || | || |:|| |: |::| : :| |: |:   : | :| |
Db   178 RFYSEDLNDLIFFPRALLN------KIVNELS------W----PLEKY 210

Qy   218 EHSVSRLQHITSSYKDYMD--TQIPALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKKWYM 275
            |:|| | |:::: | : | :| |:| | |::| :| | |:| ||::|
Db   211 EYFIQ----------MDWINQAEFIIEFLKS--------------------KKEDQ 236

Qy   276 KLKNFMVNRVFIPTKKFFNKE-----IREPSKALKEKV--STDTKDLFENKIGQGTVDFF 328
            :|:||::|| ||: | |||| || | || | ||||| ||| |||
Db   237 DLENAIDEKM----KKNLSKQQKEFILREKMKAIRSELGESNDDEEYKTKI---KDKN 288

Qy   329 NKEIRDPSKALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVYSTGAED---- 384
            |||||||||| |:| ::|||:|||||||| ||||| |:|||||| ||||| | | |
Db   289 NKKIY--PKSVRKLIANEESKLKNMSSSPDANITFRNYIDTLMKLPWRKTSIDFLDIKVA 346

Qy   385 ---LFENKIG----QGTVDF----INNEIRDP--SKALIR----KVYTEADDLF 421
               |||||:| ||||| ||||||| ||||:   :| | :|
Db   347 KKILDENHYGLKEVKERIIEFLAVLINNKNKNPDLKKSIIKLDKDHEINLNVFKESDKKF 406

Qy   422 ENKI--------GQG-------TVDFINKEIRDPSKALIRKVYSTEAD 453
            ||: |:| ||| ||:|||||||| |||| :|
Db   407 DKQITHNVPILALVGPPGTGKTSLAMAIAESINKEF---VKISLGGVKDEAE 455

---

heat shock atp-dependent proteinase [imported] - Mycoplasma pulmonis (strain UAB CTIP)

RESULT 38

G90576

C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
A;Accession: G90576
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: G90576
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-842 <KUR>
A;Cross-references: GB:AL445566; PIDN:CAC13692.1; GSPDB:GN00153

---

probable spindle pole body component, probable gamma-tubulin interacting protein, yeast

A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_5190
C;Genetic code: SGC3
A;Superfamily: ATP-dependent serine proteinase La

A:Experimental source: strain AB972
C:Genetics:
A:Gene: SGD:NIP1
A:Cross-references: SGD:S0004926; MIPS:YMR309c
A:Map position: 13R

```
Query Match     Best Local Similarity  19.5%;
                Conservative  90;  Mismatches 200; Indels 111; Gaps 22;
Matches 97;     Pred. No. 16;          Score 110.5;   DB 2;   Length 812;

Qy    6 SVGDVTKTLLAASESVDSAANAYMINSNDMSDYLSAVSD----------NFAER 48
        |:||:|                                          |  ||
Db  321 SFNDISKLLSILDQTIDT------YQVN-EFADPIDFIEDEPKEDSDGVKRIGSIFSFVER 375

Qy   49 ICSQVPKG-SNCSASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQOLEAA---- 103
Db  376 LDDEFMKSLLNIDPHSSDYLIRLRDEQSIYNLILRTQL---YFEAATLKDEHDLERALTRP 432

Qy  104 FI---------------LFKESDANPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDA 151
Db  433 FVKRLDHIYYKSENLIKIMETAAWNIIPAQFKSKF-TSKDQLDSADYVDNLIDGLSTILS 491

Qy  152 DATDIENFASRYLYMATLYY---------KTYINVDEFGAS---FFNKLSFT 191
Db  492 KQNNIAVQKRAILY--NIYYTALNKDFQTAKDMLITSQVQTNINQFDSSLQOILFNRVVVQ 549

Qy  192 TGLFGWGI--------KRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALPKF 245
Db  550 LGLSAFKLCLIEECHQILNDLLSSSHLREILGQHSLHRIS--LNSSNNASADRARQCLPY 608

Qy  246 AKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKFFNKEIREPSKALK 305
Db  609 HQHINLDLIDVVFLTCSLLIEIPRMTAFYSGIN---VNRI-----PYSPKSIRR------ 654

Qy  306 EKVSTDTKDLFENKIGQGTVDFFNK-EIRDPSKALKEKVSNDAKDLFENKIGQGTVDFIN 364
Db  655 -----------SLEHYDKLSFQGPPETLRDYLFAAKSMQKGN-WRDSVKYL- 694

Qy  365 NEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFINNEIRDPSKALIRKVYTEADDLFENK 424
Db  695 REIK--SWALLPNMETVLNSLTERVQVESLKTYFFSFPKRFYSSFSVAKL-AELFDLPENK 751

Qy  425 IGQGTVDFINKEIRDPSK 442
Db  752 VVE-VLQSVIAELEIPAK 768
```

RESULT 38
T20333
hypothetical protein D2005.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T20333
R:Wilkinson, J.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z19258
A:Molecule type: DNA
A:Residues: 1-1188 <WIL>
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Accession: T20333
A:Cross-references: EMBL:Z79752; PIDN:CAB02083.1; GSPDB:GN00019; CESP:D2005.4
A:Experimental source: clone D2005
A:Genetics:
A:Gene: CESP:D2005.4
A:Map position: 1
A:Introns: 17/2; 164/3; 199/1; 245/3; 366/3; 399/2; 414/3; 475/3; 526/3; 599/1; 679/1; 7
C:Superfamily: Caenorhabditis elegans hypothetical protein D2005.4

```
Query Match     Best Local Similarity  18.5%;
                Conservative  87;  Mismatches 183; Indels 127; Gaps 22;
Matches 90;     Pred. No. 26;          Score 110.5;   DB 2;   Length 1188;

Qy    3 PSDSVGDVTKTLLAASESVDSAANAYM--------------------INSDMSDYLSAVSDNFAERI 49
```

RESULT 39
T31679
bacitracin synthetase 3 - Bacillus licheniformis
C:Species: Bacillus licheniformis
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000
C:Accession: T31679
R:Konz, D.; Klens, A.; Schorgendorfer, K.; Marahiel, M.A.
Chem. Biol. 4, 927-937, 1997
A:Title: The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: mol
A:Reference number: T31679
A:Accession: T31679
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-6359 <KON>
A:Cross-references: EMBL:AF007865; NID:g4464275; PID:g2982196; PIDN:AAC06348.1
A:Genetics:
A:Gene: bacC
C:Superfamily: acetate--CoA ligase homology; acyl carrier protein homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:505-946/Domain: acetate--CoA ligase homology <ACL1>
F:964-1032/Domain: acyl carrier protein homology <ACP1>
F:1542-1978/Domain: acetate--CoA ligase homology <ACL2>
F:1996-2063/Domain: acyl carrier protein homology <ACP2>
F:3046-3483/Domain: acetate--CoA ligase homology <ACL3>
F:3500-3568/Domain: acyl carrier protein homology <ACP3>
F:4078-4526/Domain: acetate--CoA ligase homology <ACL4>
F:4542-4609/Domain: acyl carrier protein homology <ACP4>
F:5593-6032/Domain: acetate--CoA ligase homology <ACL5>
F:6050-6118/Domain: acyl carrier protein homology <ACP5>
F:3532/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

```
Query Match     Best Local Similarity  20.5%;
                Conservative  58;  Mismatches 173; Indels 223; Gaps 30;
Matches 117;    Pred. No. 2.2e+02;     Score 110.5;   DB 2;   Length 6359;

Db  550 PLPVVSDV-EDASSADSGKDNPESMYFQVVSTIRFGQRDGIMDRLAAFLS-VTSNTGEHI 607
Qy   50 CSQVPKGSNCSASVSAYMSRCA--KQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFIL 106
Db  608 ISLEP------------TLRFAMRCVDAKKYQTMAVDVLKFLARPQMKRMKLPNEFFEDLIVLV 659
Qy  107 FKESDANPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYM 166
Db  660 LSKYNIPDSTYSEKLY----------QFIDYLFLNLVETLRKENGIVIDEKKIVDKWANM 708
Qy  167 ATLYXKTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLP---LDIGTEH---- 219
Db  709 GHHWFSIF-----FLSHRDILKFSTTI----------AQLPWSSADIWNEEKDP 749
Qy  220 -----SVSRLQHITSSYKDYMDTQI-----PALPKFAKRFSLMVVQRLLATVAGYVDFPW 269
Db  750 SDVTSYLHLAKLPTAWPNYCNLDVIEKTVSDHIPQEVVRTQLYEAMKKYAAEMP-IDENW 808
Qy  270 YKKWYMKLKNFMVNRVFIPT-KKFFNKEIREPSKALKEKVSTDFKDLFENKIGQGTVDFF 328
Db  809 KKN---LLRHFTIKIKFLFSHKKIDKEELKEDFKQSKE---------DSKVPEKSPEVL 855
Qy  329 -NKEIRDPSKALKEKVSNDAKD----------LFENKIGQGTVDFINNEIRDP-- 370
Db  856 PGDAEKCEESEEVTTKASSDAKEGVIAVDAVAARPDIVQENAT---STDDANNEDQEKAS 912
Qy  371 -----SKALIRKVSTG----AEDLFENKIGQGTVDFINN----EIRDPSKALI-RK 412
Db  913 QVVASSVISEEVLKEESTGDCVKTSKDHMDSETAGNCSDSVINDEPVEAEDPDNVKVERK 972
Qy  413 VYTEADD 419
Db  973 EEDESID 979
```

```
Query Match     Best Local Similarity  20.5%;
                Conservative 117;  Mismatches 173; Indels 223; Gaps 30;
Matches 117;    Pred. No. 2.2e+02;     Score 110.5;   DB 2;   Length 6359;

Qy    7 VGDVTKTLLAASESVDSAANAYMINSDMSDYLSA-------VSDNFAERICSQVPKGSN 58
```

```
Db  1897  LGEIENLQLLKL-DKIDEAAVIARKDDHSDYLCAIYVSKEDWTSTEISEWLKELP----  1951
Qy    59  CSASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLITLPPDYQLEAAFILFKESDANPANST  118
Db  1952  -HYMIPAYFVR--------LDKLPLTSNDKVDRKALPAPDRHVATGAVYEA----PRNDT  1998
Qy   119  EKRFWMRFRR---------GKNHSYF----------------HDLVFNLLEKN  146
Db  1999  EAKLVDIWRDVLGAGDIGISHHFFAAGGDSIKALQIVSRLSRLGLKLEMKDLFANPRIKD  2058
Qy   147  ----------VTRDADATDIENFASRYLYMATLIYKTYTNVDEFGASF-----  184
Db  2059  LAKYVKKQSQRRNANTIVTGHAELTPIQKW-----YFA----NNKEELDHFNQSFVLFRK  2109
Qy   185  -----------FNKL-------------SFTTGLFGW------GIKRALK  204
Db  2110  GGFDESCVKKAFNKIMEQHDALRMIYEEKGGDFIQYNRSFREDLFDLDVYDVRGLDRQAE  2169
Qy   205  QI--------IRSNLPLDIGT------EHSVSRLQHITSS------YKDYMDYQI  239
Db  2170  KVYELATSIQKLSSIRKGKLVHLGIFRADEGDHLLIVIHHLVVDGVSWRILFEDFETLYS  2229
Qy   240  PAL----------PKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRV  285
Db  2230  QALKGQTLEIGYKTDSYQEFARRLKAYAHSRFLSKEAEY----W----RNIAKARV  2277
Qy   286  -FIPTKFFNKEIREPSKALKEKVSTD-TKDLFENKIGQGTVDFFNKEIRDPSKALKEKV  343
Db  2278  RFIPPKNVLKEDVYENSTTLSIKLGKEATADLLRN----TNRAYNTEIND---ILLTAL  2329
Qy   344  SNDAKDLF-ENKI----------GQGTVDFINN-EI--------RDPSKALIRKVST  380
Db  2330  LTGARDITGENLKVMMEGHGREDILEGVDITRTIGWFTTMYPVLLDAGEERALSQQIKM  2389
Qy   381  GAEDL--FENK-IGQGTV---------DFINNE  401
Db  2390  VKETLRKIPNKGIGYGLLKYMAEDPDFTNEE  2420
```

RESULT 40
HMIV
hemagglutinin precursor - influenza A virus (strain A/PR/8/34)
C;Species: Influenza A virus
C;Date:18-Dec-1981 #sequence_revision 18-Dec-1981 #text_change 28-May-1999
C;Accession: A93262; A90830; A04063
R;Winter, G.; Fields, S.; Brownlee, G.G.
Nature 292, 72-75, 1981
A;Title: Nucleotide sequence of the haemagglutinin gene of a human influenza virus H1 su
A;Reference number: A93262; MUID:82013600
A;Accession: A93262
A;Molecule type: genomic RNA
A;Residues: 1-566 <WIN>
A;Cross-references: GB:V01088; NID:g62290; PIDN:CAA24272.1; PID:g62291
R;Caton, A.J.; Brownlee, G.G.; Yewdell, J.W.; Gerhard, W.
Cell 31, 417-427, 1982
A;Title: The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subt
A;Reference number: A90830; MUID:83129356
A;Accession: A90830
A;Contents: subtype H1
A;Molecule type: genomic RNA
A;Residues: 18-145,'N',148-155,'E',157-199,'P',201-203,'E',205-207,'L',209-268,'M',270-3
A;Map position: segment 4
A;Genetics:
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thioester bond
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-343/Product: hemagglutinin HA1 chain #status predicted <HA1>
F;345-566/Product: hemagglutinin HA2 chain #status predicted <HA2>
F;555,562,565/Product: hemagglutinin HA2 chain #status predicted <HA2>
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

```
Qy   23  SAANAYMINSDMSDY--------LSAVSDNFAERICSQVPKGS-----NCSASVSAYMSRCA   71
Db  101  NSENGICYPGDFIDYEELRFQLSSVSS--FERF-EIFPKESSWPNHNTIKGVTAACSHAG  157
Qy   72  KQD-------CLTLQSLKYPLEAKYQPLTLPPDYQLEAAFILFKESDA-----NPANST  118
Db  158  KSSFYRNLLWLTEKEGSYP--------------KLKNSYVNKKGKEVLVLWGIHHPSNSK  203
Qy  119  EKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLIYMATLIYKTYTNVD  178
Db  204  DQQ--NIYQNEN-AYVSVVTSNYNRRF*TPEIAERPKVRDQAGRMNYYWTLLKPGDTIIF  259
Qy  179  EFGASFF-NKLSFTTGL-FGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSS--YKDY  234
Db  260  EANGNLIAPRYAFALSRGFGSG--------IITSNASMHECNTKCQTPLGAINSSLPFQNI  312
Qy  235  MDTQIPALPKFAKRFSLMVV---------QRLLATVAGYVDTPW---YKKWY-----  274
Db  313  HPVTIGECPKYVRSAKLRMVTGLRNIPSIOSRGLFGAIAGFIEGGWTGMIDGWYGYHHQN  372
Qy  275  -----------MKLKNFMVNRV---------FIPTKFFNKEIREPSKALKEKVSTD  311
Db  373  EQGSGYAADQKSTQNAINGITNKVNSVIEKMNIQFTAVGKEFNK-LEKRMENLNKKVDDG  431
Qy  312  TKD------LFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTVD  361
Db  432  FLDIWTYNAELLVLLEN--ERTLDFHDSNVKNLYEKVKSOLKNNAKE----IGNGCFE  483
Qy  362  F---INNEIRDPSKALIRKVSTGAEDL---FENKIGQGTVDFINNE  401
Db  484  FYHKCDNE------CMESVRNGTYDYPKYSEESKLNREKVDGVKLE  523
```

Search completed: October 19, 2002, 07:52:25
Job time : 55 secs

THIS PAGE BLANK (USPTO)